UNIVERSITATEA „AL.I.CUZA" IAȘI
FACULTATEA DE MATEMATICĂ-MECANICĂ

ELENA NENCIU

LECȚII DE STATISTICĂ MATEMATICĂ

- 1973 -

Scanned with OKEN Scanner

U N I V E R S I T A T E A „AL. I. C U Z A" I A Ş I

FACULTATEA MATEMATICĂ - MECANICĂ

Catedra de calculul probabilităţilor şi matematici aplicate

ELENA NENCIU

L E C Ţ I I

d e

S T A T I S T I C Ă M A T E M A T I C Ă

I A Ş I

- 1 9 7 2 -

Scanned with OKEN Scanner

Seredine Doina
informatică

P R E F A Ţ Ă

Ca ramură a ştiinţelor matematice, statistica matematică se ocupă de noţiunile şi metodele specifice studiului colectivităţilor statistice. Rolul statisticii matematice este de a determina, pe baza datelor furnizate de experienţă, informaţii cît mai precise asupra repartiţiei probabilistice a fenomenului de masă cercetat. In zilele noastre se manifestă un interes deosebit pentru statistica matematică şi legăturile ei cu diferite alte ştiinţe. Acest interes se materializează prin aceea că, statistica matematică este introdusă în programele de pregătire a unui număr mare de specialişti din diferite domenii de activitate.

Pornind de la faptul că statistica matematică trebuie fundamentată riguros pe teoria probabilităţilor, aceste lecţii sînt străbătute de la un capăt la altul de relaţia dintre noţiunile: mărime teoretică (probabilistică), mărime statistică (empirică) şi mărime de selecţie (funcţie de selecţie sau statistică). Aceasta constituie partea originală a lucrării. Tratarea, în acest spirit, a problemelor de bază ale statisticii matematice unidimensionale şi bidimensionale, creează condiţii pentru o abordare similară a statisticii matematice multidimensionale care nu este cuprinsă în aceste lecţii. Pentru înţelegerea lucrării cititorul trebuie să posede cunoştinţe generale de teoria probabilităţilor.

Lucrarea cuprinde următoarele capitole: Repartiţii probabilistice, Repartiţii statistice, Teoria selecţiei, Teoria estimaţiei, Verificarea ipotezelor statistice, Analiza dispersională, Corelaţia şi regresia simplă.

Lecţiile de faţă sînt destinate studenţilor de la universitate, care se profilează în domeniul statisticii matematice sau al aplicaţiilor ei. De asemenea ele sînt necesare studenţilor de la facultăţile tehnice şi economice.

Scanned with OKEN Scanner

In încheiere, aduc vii mulţumiri tov. prof. Corina Reischer pentru lectura atentă şi integrală a manuscrisului, precum şi pentru unele observaţii făcute, de care am ţinut seamă.

De asemenea îmi exprim mulţumiri colegiale tov. asistent Julien Moscovici care m-a ajutat cu multă grijă la completarea formulelor şi desenarea figurilor.

Elena Nenciu

Scanned with OKEN Scanner

CUPRINSUL

	pag.
Prefaţă	I
CAPITOLUL I. REPARTIŢII PROBABILISTICE	1
§ 1. Noţiuni de teoria probabilităţilor	2
§ 2. Repartiţii probabilistice discrete	7
§ 3. Repartiţii probabilistice continue	10
Probleme	23
CAPITOLUL II. REPARTIŢII STATISTICE	26
§ 1. Colectivitate statistică. Caracteristici	27
§ 2. Repartiţii statistice unidimensionale	28
§ 3. Medii statistice	31
§ 4. Momente statistice	35
§ 5. Funcţia de frecvenţă cumulată	42
§ 6. Cvantile şi moduri statistice	45
§ 7. Indicatorii variaţiei. Asimetrie. Exces	49
Probleme	51
CAPITOLUL III. TEORIA SELECŢIEI	54
§ 1. Fundamentele teoriei selecţiei	55
§ 2. Funcţia de repartiţie de selecţie	58
§ 3. Media de selecţie	61
§ 4. Momente iniţiale de selecţie	64
§ 5. Dispersia de selecţie	67
§ 6. Momente centrate de selecţie	71
§ 7. Teoria selecţiei dintr-o colectivitate normală unidimensională	73
Probleme	80

CAPITOLUL IV. TEORIA ESTIMAŢIEI 82

§ 1. Funcţie de estimaţie consistentă 84

§ 2. Funcţie de estimaţie absolut corectă 85

§ 3. Funcţie de estimaţie eficientă 87

§ 4. Funcţie de estimaţie corectă 99

§ 5. Funcţie de estimaţie suficientă 100

§ 6. Metode de estimare punctuală a parametrilor . . 102

§ 7. Estimarea parametrilor cu ajutorul intervalelor de încredere 109

§ 8. Intervalul de încredere pentru media unei colectivităţi normale în cazul cînd dispersia colectivităţii este cunoscută 115

§ 9. Intervalul de încredere pentru media unei colectivităţi normale în cazul cînd dispersia colectivităţii este necunoscută 117

§10. Intervalul de încredere pentru dispersia unei colectivităţi normale 119

§11. Intervalul de încredere pentru diferenţa mediilor a două colectivităţi normale în cazul cînd dispersiile colectivităţilor sînt cunoscute .120

§12. Intervalul de încredere pentru diferenţa mediilor a două colectivităţi normale în cazul cînd dispersiile colectivităţilor sînt egale cu o aceeaşi valoare necunoscută 122

§13. Intervalul de încredere pentru raportul dispersiilor a două colectivităţi normale 125

§14. Intervale de încredere pentru selecţii mari . . .127

Probleme 128

CAPITOLUL V. VERIFICAREA IPOTEZELOR STATISTICE. 133

§1. Generalităţi asupra ipotezelor statistice 134

§2. Test, funcţie de putere, funcţie caracteristică operatoare 135

§3. Cel mai puternic test. Lema Neyman-Pearson 139

Scanned with OKEN Scanner

§ 4. Teste pentru verificarea ipotezei că media unei colectivităţi normale este egală cu o valoare dată în cazul cînd dispersia colectivităţii este cunoscută 144

§ 5. Teste pentru verificarea ipotezei că media unei colectivităţi normale este egală cu o valoare dată în cazul cînd dispersia colectivităţii este necunoscută 149

§ 6. Teste pentru verificarea ipotezei că dispersia unei colectivităţi normale este egală cu o valoare dată 153

§ 7. Teste pentru verificarea ipotezei egalităţii mediilor a două colectivităţi normale în cazul cînd dispersiile colectivităţilor sînt cunoscute 157

§ 8. Teste pentru verificarea ipotezei egalităţii mediilor a două colectivităţi normale în cazul cînd dispersiile colectivităţilor sînt egale cu o aceeaşi valoare necunoscută. 161

§ 9. Teste pentru verificarea ipotezei egalităţii dispersiilor a două colectivităţi normale 165

§10. Testul raportului de verosimilităţi 169

§11. Teste cu privire la funcţia de repartiţie a colectivităţii 171

§12. Testul secvenţial al raportului probabilităţilor 175

Probleme 176

CAPITOLUL VI. ANALIZA DISPERSIONALĂ 180

§ 1. Analiza dispersională unifactorială 181

§ 2. Analiza dispersională bifactorială fără interacţiune între factori 190

Probleme 201

CAPITOLUL VII. CORELAŢIA ŞI REGRESIA SIMPLĂ 204

§ 1. Repartiţii probabilistice bidimensionale 205

§ 2. Repartiţii statistice bidimensionale 213

§ 3. Coeficientul de corelaţie teoretic. Coeficientul de contingenţă 219

§ 4. Regresia simplă teoretică 224

§ 5. Regresia teoretică liniară 227

§ 6. Raportul de corelaţie teoretic 232

§ 7. Coeficientul de corelaţie de selecţie 239

§ 8. Regresia liniară de selecţie 241

P r o b l e m e 243

A N E X A 247

Tabelul I . Valorile lui $e^{-\lambda}$ 248

Tabelul II. Repartiţia Poisson. Valorile probabilităţii

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad . . . . . \quad 249$$

Tabelul III. Repartiţia $N(0,1)$. Valorile densităţii de repartiţie $f_U(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{x^2}{2}}$ 250

Tabelul IV. Valorile funcţiei lui Laplace

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx \quad . . . . \quad 251$$

Tabelul V. Repartiţia $\chi^2(n,1)$. Valorile $\chi^2_{n,\alpha}$ în funcţie de probabilitatea $\alpha = P(X > \chi^2_{n,\alpha})$ şi numărul gradelor de libertate n 254

Tabelul VI. Repartiţia Student. Valorile funcţiei de repartiţie $F(x) = P(X < x)$ 255

Tabelul VII. Repartiţia Student. Valorile $t_{n,\alpha}$ în funcţie de probabilitatea $\alpha = P(|X| > t_{n,\alpha})$ şi numărul gradelor de libertate n 257

Tabelul VIII. Repartiţia Snedecor. Valorile $F_{m,n};0,05$ în funcţie de probabilitatea $\alpha = P(X > F_{m,n};0,05) = 0,05$ şi numerele gradelor de libertate m şi n 258

Tabelul IX. Repartiţia Snedecor. Valorile $F_{m,n};0,10$ în funcţie de probabilitatea $\alpha = P(X > F_{m,n};0,10) = 0,10$ şi numerele gradelor de libertate m şi n 260

Tabelul X. Valorile funcţiei lui Kolmogorov

$$K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$$ 262

B i b l i o g r a f i e 264

Capitolul I

REPARTIȚII PROBABILISTICE

Scanned with OKEN Scanner

§ 1. NOŢIUNI DE TEORIA PROBABILITĂŢILOR

Fie $X(\omega)$ o variabilă aleatoare definită pe cîmpul de evenimente $\{\Omega, \mathcal{K}\}$. A cunoaşte legea de probabilitate sau repartiţia probabilistică a variabilei aleatoare X, înseamnă a şti cum se calculează $P(\omega \in \Omega : X(\omega) \in B)$, pentru orice mulţime borelianǎ B, a lui R. Repartiţia probabilistică a unei variabile aleatoare este complet determinată de funcţia sa de repartiţie.

Definiţia 1. Se numeşte funcţie de repartiţie a variabilei aleatoare $X(\omega)$, aplicaţia $F : R \longrightarrow R$, definită prin

$$F(x) = P(\omega \in \Omega : X(\omega) < x)$$

pentru orice $x \in R$.

Evident, $F(x) = P(X^{-1}(-\infty, x))$ pentru orice $x \in R$.

Definiţia 2. Se spune că repartiţia probabilistică a variabilei aleatoare X este discretă dacă funcţia sa de repartiţie este de forma

$$F(x) = \sum_{\{k : x_k < x\}} p_k$$

unde x_k sînt puncte de discontinuitate pentru $F(x)$, p_k sînt salturile lui $F(x)$ în x_k, $(k \in I)$, iar I este o mulţime cel mult numărabilă.

Definiţia 3. Se spune că repartiţia probabilistică a variabilei aleatoare X este continuă dacă funcţia sa de repartiţie este de forma

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

unde $f : R \longrightarrow R$, măsurabilă Borel.

Funcţia $f(x)$ se numeşte densitate de repartiţie.

Definiţia 4. Se numeşte valoare medie (medie) a variabilei aleatoare X , numărul

$$M(X) = \int_{-\infty}^{+\infty} x \, dF(x)$$

dacă integrala din membrul drept există.

Definiţia 5. Se numeşte moment iniţial de ordin k al variabilei aleatoare X, $(k = 0,1,2,\ldots)$, numărul

$$\alpha_k(X) = M(X^k)$$

dacă media din membrul drept există.

Evident, $\alpha_1(X) = M(X)$.

Definiţia 6. Se numeşte dispersie a variabilei aleatoare X, numărul $D^2(X) = M\left[(X - M(X))^2\right]$.

dacă media din membrul drept există.

Definiţia 7. Se numeşte abatere medie pătratică a variabilei aleatoare X, numărul

$$D(X) = \sqrt{D^2(X)}$$

dacă $D^2(X)$ există.

Definiţia 8. Se numeşte covarianţă (corelaţie) a variabilelor aleatoare X şi Y, numărul

$$\text{cov.}(X,Y) = M\left[(X - M(X))(Y - M(Y))\right]$$

dacă media din membrul drept există.

Definiţia 9. Se numeşte coeficient de corelaţie a variabilelor aleatoare X şi Y, numărul

$$\rho(X,Y) = \frac{\text{cov.}(X,Y)}{D(X)D(Y)}$$

dacă $D(X)D(Y) \neq 0$.

Definiţia 10. Se numeşte variabilă aleatoare normată a variabilei aleatoare X, variabila aleatoare

$$Z(\omega) = \frac{X(\omega) - M(X)}{D(X)}$$

Evident, $M(Z) = 0$ şi $D^2(Z) = 1$, oricare ar fi variabila aleatoare X.

Definiţia 11. Se numeşte moment absolut de ordin r al variabilei aleatoare X, $(0 < r \in R)$, numărul

$$\beta_r(X) = M(|X|^r)$$

dacă media din membrul drept există.

Definiţia 12. Se numeşte moment centrat de ordin k al variabilei aleatoare X, $(k = 0,1,2,\ldots)$, numărul

$$\mu_k(X) = M\left[(X - M(X))^k\right]$$

dacă media din membrul drept există.

Evident, $\mu_2(X) = D^2(X)$.

Intre momentele iniţiale şi centrate ale variabilei aleatoare X, au loc formulele de legătură

$$\mu_k(X) = \sum_{i=0}^{k} (-1)^i C_k^i \alpha_{k-i}(X)\alpha_1^i(X)$$

$$\alpha_k(X) = \sum_{i=0}^{k} C_k^i \mu_{k-i}(X)\alpha_1^i(X)$$

pentru $k = 1,2,3,\ldots$

Inegalitatea lui Cebîşev. Dacă variabila aleatoare X admite medie şi dispersie finite, atunci pentru orice $a \in R$, $a > 0$, are loc

$$P(\omega \in \Omega : |X - M(X)| \geqslant a) \leqslant \frac{D^2(X)}{a^2} .$$

Inegalitatea lui Schwarz. Dacă există şi sînt finite $\beta_2(X)$ şi $\beta_2(Y)$, atunci există şi este finit $\beta_1(XY)$ şi are loc

$$\beta_1(XY) \leqslant \sqrt{\beta_2(X)\beta_2(Y)}$$

Observaţie. Inegalitatea lui Schwarz este echivalentă cu inegalitatea

$$M^2(|XY|) \leqslant M(X^2)M(Y^2)$$

Definiţia 13. Se numeşte funcţie caracteristică a variabilei aleatoare X, aplicaţia $\varphi : R \longrightarrow C$, definită prin

$$\varphi(t) = M(e^{itX})$$

unde $i = \sqrt{-1}$.

Funcţia caracteristică există pentru orice variabilă aleatoare.

Fie $N^* = \{1,2,3,\ldots\}$ şi $(X_n)_{n\in N^*}$ un şir de variabile aleatoare definite pe cîmpul de probabilitate $\{\Omega, \mathcal{K}, P\}$.

Definiţia 14. Se spune că şirul de variabile aleatoare $(X_n)_{n\in N^*}$ converge în probabilitate (slab) către variabila aleatoare X, dacă pentru orice $\varepsilon > 0$, are loc

$$\lim_{n\to\infty} P(\omega\in\Omega : |X_n - X| \geqslant \varepsilon) = 0 .$$

Definiţia 15. Se spune că şirul de variabile aleatoare $(X_n)_{n\in N^*}$ converge aproape sigur (tare) către variabila aleatoare X, dacă pentru orice $\varepsilon > 0$, are loc

$$\lim_{k\to\infty} P(\omega\in\Omega : \sup_{n\geqslant k} |X_n - X| \geqslant \varepsilon) = 0$$

Definiţia 16. Se spune că şirul de variabile aleatoare $(X_n)_{n\in N^*}$ converge în medie de ordin r , $(0 < r \in R)$, către variabila aleatoare X, dacă în ipoteza că există şi sînt finite $\beta_r(X_n)$, $n \in N^*$ şi $\beta_r(X)$, are loc

$$\lim_{n\to\infty} \beta_r(X_n - X) = 0 .$$

Definiţia 17. Se spune că şirul de variabile aleatoare $(X_n)_{n\in N^*}$ converge în repartiţie către variabila aleatoare X, dacă are loc

$$\lim_{n\to\infty} F_n(x) = F(x)$$

în orice $x \in R$, punct de continuitate pentru $F(x)$, unde $F(x)$ este funcţia de repartiţie a lui X, iar $F_n(x)$ este funcţia de repartiţie a lui X_n.

Definiţia 18. Se spune că variabilele aleatoare $X(\omega)$ şi $Y(\omega)$ sînt egale aproape sigur, dacă are loc

$$P(\omega\in\Omega : X(\omega) = Y(\omega)) = 1 .$$

Teorema lui Bernoulli. Dacă $f_n(\omega)$ este frecvenţa relativă, în n probe independente, a unui eveniment ce se produce cu aceeaşi probabilitate p în fiecare probă, atunci $(f_n)_{n\in N^*}$ converge în probabilitate către p .

Teorema lui Hincin. Dacă variabilele aleatoare ale şirului $(X_n)_{n\in N^*}$, satisfac condiţiile :

i) sînt independente două cîte două

ii) au aceeaşi funcţie de repartiţie

iii) $M(X_n) = c < +\infty$, $n \in N^*$

atunci $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)_{n\in N^*}$ converge în probabilitate către c .

Teorema lui Borel. Dacă $f_n(\omega)$ este frecvenţa relativă, în n probe independente, a unui eveniment ce se produce cu aceeaşi probabilitate p în fiecare probă, atunci $(f_n)_{n\in N^*}$ converge aproape sigur către p .

Fie $(Y_n)_{n\in N^*}$ un şir de variabile aleatoare, independente, ce admit medii şi dispersii finite. Notăm

$$M(Y_i) = m_i \ , \ D^2(Y_i) = D_i^2 \ , \quad i \in N^*$$

$$S_n = \sqrt{D_1^2 + \ldots + D_n^2} \ , \ Z_n(\omega) = \frac{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} m_i}{S_n} \ , \quad n \in N^*$$

O teoremă limită centrală stabileşte condiţiile în care şirul de variabile aleatoare $(Z_n)_{n\in N^*}$ converge în repartiţie către o variabilă aleatoare ce urmează repartiţia N(0,1). Una din teoremele limită centrală ce are loc în condiţii foarte generale este

Teorema lui Leapunov. Fie $(Y_n)_{n\in N^*}$ un şir de variabile aleatoare independente, pentru care există şi sînt finite :

$$M(Y_i) = m_i \ , \ D^2(Y_i) = D_i^2 \ , \ M(|Y_i - m_i|^3) = H_i^3 \ , \quad i \in N^*$$

Fie

$$S_n=\sqrt{D_1^2+\ldots+D_n^2}\ ,\quad K_n=\sqrt[3]{H_1^3+\ldots+H_n^3},\quad Z_n=\frac{\sum_{i=1}^{n}Y_i-\sum_{i=1}^{n}m_i}{S_n}\ ,\quad n\in N^*$$

Dacă

$$\lim_{n\to\infty}\frac{K_n}{S_n}=0$$

atunci $(Z_n)_{n\in N^*}$ converge în repartiţie către o variabilă aleatoare ce urmează repartiţia $N(0,1)$.

Evident, teorema lui Leapunov rămîne valabilă şi în cazul cînd variabilele aleatoare ale şirului $(Y_n)_{n\in N^*}$ urmează aceeaşi repartiţie.

§ 2. REPARTIŢII PROBABILISTICE DISCRETE.

O repartiţie discretă a unei variabile aleatoare X este complet determinată de tabloul de repartiţie care are forma

$$X:\begin{bmatrix}x_i\\p_i\end{bmatrix}_{i\in I}$$

cu $\sum_{i\in I}p_i=1$, unde x_i sînt valorile posibile ale variabilei aleatoare X, $p_i=P(\omega\in\Omega:X(\omega)=x_i)$, $i\in I$, iar I este o mulţime cel mult numărabilă.

Repartiţia unei variabile aleatoare ce poate lua numai două valori este dată de

$$X:\begin{bmatrix}0&1\\q&p\end{bmatrix}$$

cu $p+q=1$.

Dacă în fiecare probă a unei experienţe un eveniment oarecare A are probabilitatea de realizare p şi probabilitatea de nerealizare q, atunci numărul de realizări ale evenimentului A într-o singură probă, reprezintă o variabilă aleatoare ce poate lua numai două valori.

Repartiţia uniformă este dată de

$$X : \begin{bmatrix} k \\ \frac{1}{n} \end{bmatrix}_{1 \leq k \leq n}$$

Repartiţia binominală sau repartiţia $B(n,p)$ este dată de

$$X : \begin{bmatrix} k \\ C_n^k p^k q^{n-k} \end{bmatrix}_{0 \leq k \leq n}$$

unde $p + q = 1$.

Evident, dacă X urmează repartiţia $B(n,p)$, atunci

$X = \sum_{i=1}^{n} X_i$, unde $X_i : \begin{bmatrix} 0 & 1 \\ q & p \end{bmatrix}$, $1 \leq i \leq n$

cu $p + q = 1$ şi variabilele aleatoare X_i, $1 \leq i \leq n$, sînt independente. Un exemplu de variabilă aleatoare ce urmează repartiţia $B(n,p)$ îl constituie frecvenţa absolută, în n probe independente, a unui eveniment ce se produce cu aceeaşi probabilitate p în fiecare probă.

Teorema 1. Dacă X urmează repartiţia $B(n,p)$, atunci

$$M(X) = np, \quad D^2(X) = npq .$$

Teorema 2. Funcţia caracteristică a unei variabile aleatoare X ce urmează repartiţia $B(n,p)$ este

$$\varphi(t) = (pe^{it} + q)^n$$

Teorema 3. Dacă X_1 şi X_2 sînt variabile aleatoare independente, urmînd repartiţiile $B(n_1,p)$, respectiv $B(n_2,p)$, atunci variabila aleatoare $X_1 + X_2$ urmează repartiţia $B(n_1 + n_2, p)$.

Repartiţia Poisson de parametru λ este dată de

$$X : \begin{bmatrix} k \\ e^{-\lambda} \frac{\lambda^k}{k!} \end{bmatrix}, \quad k = 0,1,2,\ldots$$

unde $\lambda > 0$.

Teorema 4. Dacă $n \to \infty$ şi $p \to 0$ astfel încît $np=\lambda$, atunci

$$\lim_{n\to\infty} C_n^k p^k q^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!} .$$

Teorema 5. Dacă X urmează repartiţia Poisson de parametru λ, atunci

$$M(X) = D^2(X) = \lambda .$$

Teorema 6. Funcţia caracteristică a unei variabile aleatoare X ce urmează repartiţia Poisson de parametru λ este

$$\varphi(t) = e^{\lambda(e^{it}-1)} .$$

Teorema 7. Dacă X_1 şi X_2 sînt variabile aleatoare independente, urmînd repartiţiile Poisson de parametru λ_1, respectiv λ_2, atunci variabila aleatoare $X_1 + X_2$ urmează repartiţia Poisson de parametru $\lambda_1 + \lambda_2$.

Dacă X urmează repartiţia Poisson de parametru λ, atunci valorile probabilităţii

$$P(\omega\in\Omega;\ X(\omega) = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

pentru diferite valori ale lui λ şi k se află în tabelul II din anexă.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia Poisson de parametru $\lambda = 2,5$. Să se determine probabilitatea ca variabila aleatoare X să ia valoarea 8.

In tabelul II din anexă, la intersecţia liniei $k = 8$ cu coloana $\lambda = 2,5$ se află

$$P(\omega\in\Omega;\ X(\omega) = 8) = 0,0031 .$$

§ 3. REPARTIŢII PROBABILISTICE CONTINUE.

O repartiţie probabilistică continuă a unei variabile aleatoare X este complet determinată de densitatea de repartiţie f(x), care satisface condiţiile:

i) $f(x) \geq 0$ pentru orice $x \in R$

ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Teorema 1. Fie f(x,y) densitatea de repartiţie comună a variabilelor aleatoare X şi Y, g(u,v) densitatea de repartiţie comună a variabilelor aleatoare U şi V, iar

$$\tau : \begin{cases} x = \varphi(u,v) \\ y = \psi(u,v) \end{cases}$$

o transformare, unde φ şi ψ sînt aplicaţii biunivoce şi diferenţiabile ale lui R^2 în R . Atunci, densitatea de repartiţie comună a variabilelor aleatoare U şi V este

$$g(u,v) = f\left[\varphi(u,v),\ \psi(u,v)\right].|J|$$

unde J este iacobianul transformării τ .

Repartiţia normală standard sau repartiţia N(0,1) este dată de densitatea de repartiţie

$$f_U(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Graficul lui $f_U(x)$ este cel din figura 1.

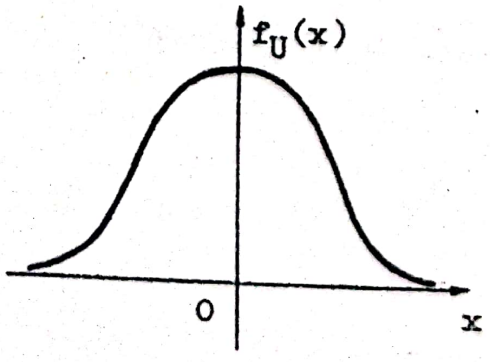


Fig.1.

Valorile lui $f_U(x)$ pentru x>0 se află în tabelul III din anexă. Datorită faptului că $f_U(-x) = f_U(x)$, pentru orice $x \in R$, sînt suficiente tabelele numai pentru $x > 0$.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia

$N(0,1)$. Să se determine valoarea densității de repartiție $f_U(x)$ în punctul $x = 2,9$.

In tabelul III din anexă, la intersecția liniei $x = 2,9$ cu coloana lui $f_U(x)$, se află

$$f_U(2,9) = 0,00595$$

Teorema 2. Dacă X urmează repartiția $N(0,1)$, atunci

$$M(X) = 0 , \quad D^2(X) = 1 .$$

Teorema 3. Funcția caracteristică a unei variabile aleatoare X ce urmează repartiția $N(0,1)$ este

$$\varphi(t) = e^{-\frac{t^2}{2}}$$

Funcția de repartiție a variabilei aleatoare ce urmează repartiția $N(0,1)$ este

$$F_U(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{v^2}{2}} dv .$$

Pentru calculul valorilor lui $F_U(x)$ se folosește funcția lui Laplace, definită prin

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{x^2}{2}} dx .$$

Proprietățile pe care le posedă $\Phi(z)$ sînt :

i) $\Phi(0) = 0$

ii) $\Phi(-z) = -\Phi(z)$ oricare ar fi $z \in R$

iii) $\Phi(+\infty) = 0,5$

Valorile lui $\Phi(z)$ pentru $z > 0$ se află în tabelul IV din anexă.

Intre $F_U(x)$ și $\Phi(x)$ există relația

$$F_U(x) = 0,5 + \Phi(x)$$

pentru orice $x \in R$.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiția $N(0,1)$. Să se determine valoarea funcției de repartiție $F_U(x)$ în punctul $x = -1,25$.

Folosind relaţia dintre $F_U(x)$ şi $\Phi(x)$ precum şi proprietăţile lui $\Phi(x)$ se obţine

$$F_U(-1,25) = 0,5 + \Phi(-1,25) = 0,5 - \Phi(1,25)$$

In tabelul IV din anexă, la intersecţia liniei $z = 1,25$ cu coloana lui $\Phi(z)$, se află $\Phi(1,25) = 0,39435$. Prin urmare $F_U(-1,25) = 0,5 - 0,39435 = 0,10565$.

In legătură cu o variabilă aleatoare X ce urmează repartiţia $N(0,1)$ se pun următoarele probleme :

i) Date numerele $a,b \in R$, să se determine

$$P(\omega \in \Omega : a < X(\omega) < b)$$

După cum se ştie

$$P(\omega \in \Omega : a < X(\omega) < b) = F_U(b) - F_U(a) .$$

Ţinînd seama de relaţia dintre $F_U(x)$ şi $\Phi(x)$, ultima relaţie dev

$$P(\omega \in \Omega : a < X(\omega) < b) = \Phi(b) - \Phi(a) .$$

In cazu rticular, $a = -u$, $b = u$, se obţine

$$P(\omega \in \Omega : -u < X(\omega) < u) = 2\Phi(u)$$

Prin urmare, probabilitatea necunoscută se determină cu ajutorul funcţiei lui Laplace, folosind tabelul IV din anexă.

ii) Dată probabilitatea δ , să se determine $u > 0$, astfel încît $P(\omega \in \Omega : -u < X(\omega) < u) = \delta$

Ţinînd seama de rezolvarea problemei i), avem $2\Phi(u) = \delta$. Valoarea necunoscută u se determină cu ajutorul funcţiei lui Laplace, pe baza relaţiei $\Phi(u) = \frac{\delta}{2}$ folosind tabelul IV din anexă.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia $N(0,1)$. Se cere :

i) $P(\omega \in \Omega : -2,45 < X(\omega) < 1,27)$

ii) valoarea u, încît

$$P(\omega \in \Omega : -u < X(\omega) < u) = 0,98222 .$$

Ţinînd seama de cele spuse mai înainte, avem

i) $P(\omega\in\Omega : -2{,}45 < X(\omega) < 1{,}27) = \Phi(1{,}27) - \Phi(-2{,}45) =$
$= \Phi(1{,}27) + \Phi(2{,}45) = 0{,}39796 + 0{,}49286 = 0{,}89082$

ii) $P(\omega\in\Omega : -u < X(\omega) < u) = 2\Phi(u) = 0{,}98222$ de unde
$\Phi(u) = 0{,}49111$.

Folosind tabelul IV din anexă se obţine $u = 2{,}37$.

Repartiţia normală generală sau repartiţia $N(m,\sigma^2)$ este dată de densitatea de repartiţie

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

unde $m\in R$, $\sigma > 0$.

Graficul lui $f(x)$ este cel din figura 2.

Teorema 4. Dacă variabila aleatoare X urmează repartiţia $N(m,\sigma^2)$, atunci

$M(X) = m$, $D^2(X) = \sigma^2$.

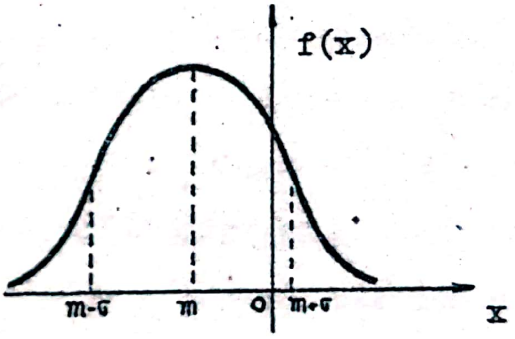


Fig.2.

Teorema 5. Funcţia caracteristică a unei variabile aleatoare X ce urmează repartiţia $N(m,\sigma^2)$ este

$$\varphi(t) = e^{mit - \frac{t^2\sigma^2}{2}}$$

Teorema 6. Dacă X urmează repartiţia $N(m,\sigma^2)$, atunci variabila aleatoare formată

$$Z = \frac{X - m}{\sigma}$$

urmează repartiţia $N(0,1)$.

Demonstraţie. Se foloseşte metoda funcţiei caracteristice. Conform teoremei 5, funcţia caracteristică a lui X este

$$\varphi_X(t) = e^{mit - \frac{t^2\sigma^2}{2}}$$

Funcţia caracteristică a lui Z va fi

$$\Psi(t) = M(e^{itZ}) = M(e^{it\frac{X-m}{\sigma}}) = e^{-mi\frac{t}{\sigma}} M(e^{i\frac{t}{\sigma}X}) =$$

$$= e^{-mi\frac{t}{\sigma}} \varphi_X(\frac{t}{\sigma}) = e^{-\frac{t^2}{2}}$$

care, conform teoremei 3, este tocmai funcţia caracteristică a repartiţiei N(0,1).

Teorema 7. Dacă variabilele aleatoare X_i, $1 \leq i \leq n$, sînt independente şi urmează repartiţiile $N(m_i, \sigma_i^2)$, $1 \leq i \leq n$, iar $a_i \in R$, $1 \leq i \leq n$, atunci variabila aleatoare $X = \sum_{i=1}^{n} a_i X_i$ urmează repartiţia $N(\sum_{i=1}^{n} a_i m_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$.

Demonstraţie. Se foloseşte metoda funcţiei caracteristice.

Dacă $a_1 = -1$, $a_2 = 1$, $a_3 = a_4 = \ldots = a_n = 0$, atunci potrivit teoremei 7, variabila aleatoare $X_1 - X_2$ urmează repartiţia $N(m_1 - m_2, \sigma_1^2 + \sigma_2^2)$.

Funcţia de repartiţie a unei variabile aleatoare ce urmează repartiţia $N(m, \sigma^2)$ este

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-m)^2}{2\sigma^2}} dv .$$

Intre F(x) şi funcţia lui Laplace $\Phi(x)$, există relaţia

$$F(x) = 0{,}5 + \Phi(\frac{x-m}{\sigma})$$

pentru orice $x \in R$. Dacă sînt date numerele $a, b \in R$, iar X urmează repartiţia $N(m, \sigma^2)$, atunci

$$P(\omega \in \Omega : a < X(\omega) < b) = \Phi(\frac{b-m}{\sigma}) - \Phi(\frac{a-m}{\sigma}) .$$

Repartiţia hi-pătrat cu n grade de libertate şi parametru σ sau repartiţia $\chi^2(n, \sigma)$ este dată de densitatea de repartiţie

$$f_{\chi^2}(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} e^{-\frac{x}{2\sigma^2}}}{2^{\frac{n}{2}} \sigma^n \Gamma\left(\frac{n}{2}\right)} & \text{dacă } x > 0 \\ 0 & \text{dacă } x \leq 0 \end{cases}$$

unde $\sigma > 0$ şi $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$, $(a > 0)$.

Pentru $n > 2$ şi $\sigma = 1$, graficul densităţii de repartiţie $f_{\chi^2}(x)$ este cel din figura 3.

Valorile $\chi^2_{n,\alpha}$ ale unei variabile aleatoare X ce urmează repartiţia $\chi^2(n,1)$, în funcţie de probabilitatea $\alpha = P(X > \chi^2_{n,\alpha})$ şi numărul gradelor de libertate n, sînt date în tabelul V din anexă.

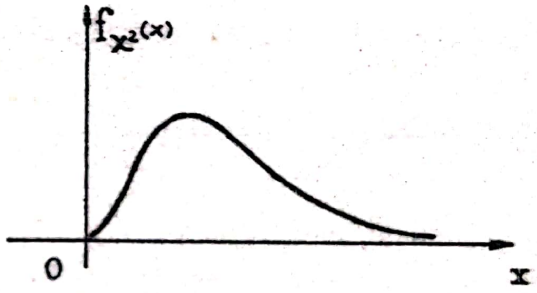


Fig.3.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia $\chi^2(10,1)$. Să se determine valoarea $\chi^2_{10;0,05}$.

In tabelul V din anexă, la intersecţia liniei $n = 10$ cu coloana $\alpha = 0,05$, se află $\chi^2_{10;0,05} = 18,307$.

Teorema 8. Funcţia caracteristică a unei variabile aleatoare ce urmează repartiţia $\chi^2(n,\sigma)$ este

$$\varphi(t) = (1 - 2it\sigma^2)^{-\frac{n}{2}}$$

Teorema 9. Dacă variabilele aleatoare X_i, $1 \leq i \leq n$, sînt independente şi urmează repartiţiile $\chi^2(k_i, \sigma)$, $1 \leq i \leq n$, atunci variabila aleatoare $X = \sum_{i=1}^{n} X_i$ urmează repartiţia $\chi^2(\sum_{i=1}^{n} k_i, \sigma)$.

Demonstraţie. Se foloseşte metoda funcţiei caracteristice.

Teorema 10. Dacă variabilele aleatoare X_k, $1 \leq k \leq n$, sînt independente şi urmează aceeaşi repartiţie $N(0, \sigma^2)$, atunci variabila aleatoare $\chi^2 = \sum_{k=1}^{n} X_k^2$ urmează repartiţia $\chi^2(n,\sigma)$.

Demonstraţie. Fie $F(x)$ şi $f(x)$, funcţia de repartiţie, respectiv densitatea de repartiţie, a repartiţiei $N(0, \sigma^2)$. Prin urmare

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

Fie $G(y)$ şi $g(y)$, funcţia de repartiţie, respectiv densitatea de repartiţie a variabilei aleatoare X_k^2, $1 \leq k \leq n$. Avem

$$G(y) = P(X_k^2 < y) = \begin{cases} 0 & \text{dacă } y \leq 0 \\ P(-\sqrt{y} < X_k < \sqrt{y}) & \text{dacă } y > 0 \end{cases} =$$

$$= \begin{cases} 0 & \text{dacă } y \leq 0 \\ F(\sqrt{y}) - F(-\sqrt{y}) & \text{dacă } y > 0 \end{cases}$$

Derivînd ultima relaţie, se obţine

$$G'(y) = g(y) = \begin{cases} 0 & \text{dacă } y \leq 0 \\ \frac{1}{2\sqrt{y}}\left[f(\sqrt{y}) + f(-\sqrt{y})\right] & \text{dacă } y > 0 \end{cases} =$$

$$= \begin{cases} 0 & \text{dacă } y \leq 0 \\ \dfrac{y^{-\frac{1}{2}} e^{-\frac{y}{2\sigma^2}}}{\sigma\sqrt{2\pi}} & \text{dacă } y > 0 \end{cases}$$

Funcţia caracteristică a variabilei aleatoare X_k^2 este

$$\varphi_{X_k^2}(t) = M(e^{itX_k^2}) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} y^{-\frac{1}{2}} e^{-\frac{1-2it\sigma^2}{2\sigma^2}} dy = (1-2it\sigma^2)^{-\frac{1}{2}}.$$

Din independenţa variabilelor aleatoare X_k, $1 \leq k \leq n$, rezultă că funcţia caracteristică a variabilei aleatoare

$\chi^2 = \sum_{k=1}^{n} X_k^2$ va fi

$$\phi(t) = M(e^{it\sum_{k=1}^{n} X_k^2}) = \prod_{k=1}^{n} M(e^{itX_k^2}) = \prod_{k=1}^{n} \varphi_{X_k^2}(t) =$$

$$= (1 - 2\,it\,\sigma^2)^{-\frac{n}{2}}$$

care, conform teoremei 8, este tocmai funcţia caracteristică a unei variabile aleatoare ce urmează repartiţia $\chi^2(n, \sigma)$.

Teorema 11. Fie

$$Q(X_1, X_2, \ldots, X_n) = \sum_{k=1}^{n} X_k^2 - \sum_{j=1}^{s} \ell_j^2(X_1, X_2, \ldots, X_n), \quad (s < n),$$

unde X_k, $1 \leqslant k \leqslant n$, sînt variabile aleatoare independente urmînd aceeaşi repartiţie $N(0,1)$, iar

$$\ell_j(X_1, X_2, \ldots, X_n) = \sum_{k=1}^{n} a_{jk} X_k, \quad 1 \leqslant j \leqslant s,$$

sînt forme liniare în X_k, $1 \leqslant k \leqslant n$, independente şi urmînd de asemenea repartiţia $N(0,1)$. Atunci, variabila aleatoare $Q(X_1, X_2, \ldots, X_n)$ urmează repartiţia $\chi^2(n - s, 1)$.

Demonstraţie. A se vedea Onicescu O., Mihoc Gh. [12].

Observaţia 1. Teorema 11 se poate demonstra şi pentru cazul cînd variabilele aleatoare X_k, $1 \leqslant k \leqslant n$, urmează repartiţia $N(0, \sigma^2)$.

Teorema 12. (Cochran). Fie X_k, $1 \leqslant k \leqslant n$, variabile aleatoare independente urmînd aceeaşi repartiţie $N(0,1)$ şi $Q_1, Q_2, \ldots, Q_s$ forme pătratice în X_k, $1 \leqslant k \leqslant n$, respectiv de rangul $k_1, k_2, \ldots, k_s$. Dacă

$$\sum_{h=1}^{s} Q_h(X_1, X_2, \ldots, X_n) = \sum_{k=1}^{n} X_k^2$$

atunci condiţia necesară şi suficientă ca variabilele aleatoare

$Q_1, Q_2, \ldots, Q_s$ să fie independente este $k_1 + k_2 + \ldots + k_s = n$. Repartiţia probabilistică a lui Q_b, este repartiţia $\chi^2(k_b, 1)$, $1 \leq b \leq s$.

Demonstraţie. A se vedea Onicescu O., Mihoc Gh., [12].

Observaţia 2. Teorema 12 se poate demonstra şi în cazul cînd variabilele aleatoare X_k, $1 \leq k \leq n$, urmează repartiţia $N(0, \sigma^2)$.

Repartiţia Student cu n grade de libertate sau repartiţia t este dată de densitatea de repartiţie

$$f_t(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} .$$

Graficul lui $f_t(x)$ este cel din figura 4.

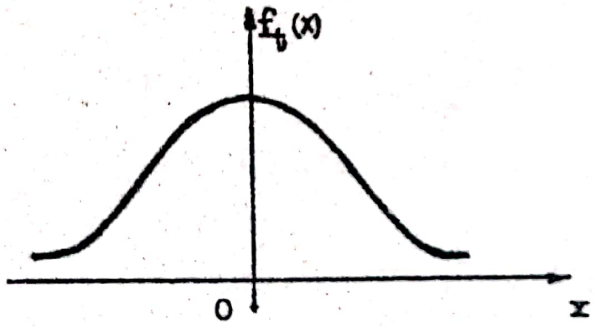


Fig.4.

Valorile $t_{n,\alpha}$ ale unei variabile aleatoare X ce urmează repartiţia Student, în funcţie de probabilitatea $\alpha = P(|X| > t_{n,\alpha})$ şi numărul gradelor de libertate n, sînt date în tabelul VII din anexă.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia Student cu 9 grade de libertate. Să se determine valoarea $t_{9;,0,05}$.

In tabelul VII din anexă, la intersecţia liniei $n = 9$ cu coloana $\alpha = 0,05$ se află $t_{9;0,05} = 2,262$.

Teorema 13. Dacă variabilele aleatoare U şi V sînt independente şi urmează repartiţiile $N(0, \sigma^2)$, respectiv $\chi^2(n, \sigma)$, atunci variabila aleatoare

$$t = \frac{U}{\sqrt{\frac{V}{n}}}$$

urmează repartiţia Student cu n grade de libertate.

Demonstraţie. Fie

$$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

densitatea de repartiţie a lui U ,

$$f_2(y) = \begin{cases} \dfrac{y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}}{2^{\frac{n}{2}}\sigma^n \Gamma(\frac{n}{2})} & \text{dacă } y > 0 \\ 0 & \text{dacă } y \leq 0 \end{cases}$$

densitatea de repartiţie a lui V şi $f(x,y)$ densitatea de repartiţie comună a variabilelor aleatoare U şi V. Din independenţa variabilelor aleatoare U şi V , rezultă

$$f(x,y) = f_1(x)f_2(y) = \frac{y^{\frac{n}{2}-1} e^{-\frac{x^2+y}{2\sigma^2}}}{2^{\frac{n+1}{2}}\sigma^{n+1}\sqrt{\pi}\,\Gamma(\frac{n}{2})}$$

Fie transformarea

$$\tau : \begin{cases} x = t\sqrt{\frac{v}{n}} \\ y = v \end{cases}$$

al cărei iacobian este $J = \sqrt{\frac{v}{n}}$. Conform teoremei 1, densitatea de repartiţie comună a variabilelor aleatoare T şi V este

$$g(t,v) = \frac{v^{\frac{n}{2}-1} e^{-\frac{v}{2\sigma^2}(1+\frac{t^2}{n})}}{2^{\frac{n+1}{2}}\sigma^{n+1}\sqrt{\pi}\,\Gamma(\frac{n}{2})}\sqrt{\frac{v}{n}}$$

Densitatea de repartiţie a lui T se obţine integrînd $g(t,v)$ în raport cu v . Avem

$$h(t) = \int_0^{+\infty} g(t,v)dv = \frac{1}{2^{\frac{n+1}{2}} \sigma^n \sqrt{\pi}\, \Gamma(\frac{n}{2})} \int_0^{+\infty} v^{\frac{n}{2}-1} e^{-\frac{v}{2\sigma^2}(1+\frac{t^2}{n})} \sqrt{\frac{v}{n}}\, dv$$

In urma schimbării de variabilă

$$z = \frac{v}{2\sigma^2}(1 + \frac{t^2}{n})$$

se obţine

$$h(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}(1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$$

care constituie densitatea de repartiţie Student cu n grade de libertate.

Observaţia 3. Folosind teoremele 10 şi 13 se poate demonstra că dacă variabilele aleatoare $X_o, X_1, X_2, \ldots, X_n$ sînt independente şi urmează aceeaşi repartiţie N(0,1), atunci variabila aleatoare

$$t = \frac{X_o}{\sqrt{\frac{X_1^2+\ldots+X_n^2}{n}}}$$

urmează repartiţia Student cu n grade de libertate.

Repartiţia Snedecor sau repartiţia F cu m şi n grade de libertate este dată de densitatea de repartiţie

$$f_F(x) = \begin{cases} \dfrac{(\frac{m}{n})^{\frac{m}{2}}\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} x^{\frac{m}{2}-1}(1+\frac{m}{n}x)^{-\frac{m+n}{2}} & \text{dacă } x > 0 \\ 0 & \text{dacă } x \leqslant 0 \end{cases}$$

Graficul lui $f_F(x)$, pentru $m > 2$, este cel din figura 5. Valorile $F_{m,n;\alpha}$ ale unei variabile aleatoare X ce urmează repartiţia Snedecor, în funcţie de probabilitatea $\alpha = P(X > F_{m,n;\alpha})$ şi numerele gradelor de libertate m şi n sînt date în tabelul VIII pentru $\alpha = 0{,}05$ şi în tabelul IX pentru $\alpha = 0{,}10$.

Observaţia 4. Există tabele de valori $F_{m,n;\alpha}$ numai pentru $\alpha < 0,5$. Valorile $F_{m,n;\alpha}$ cu $\alpha > 0,5$ se calculează pe baza relaţiei

$$F_{m,n;\alpha} = \frac{1}{F_{n,m;1-\alpha}}$$

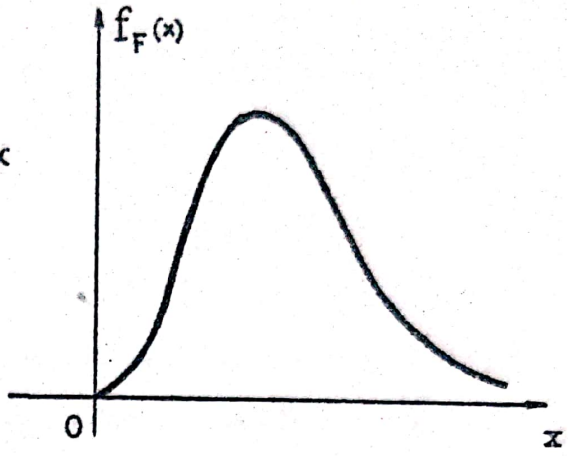


Fig. 5.

Exemplu. Fie X o variabilă aleatoare ce urmează repartiţia Snedecor cu 10 şi 4 grade de libertate. Să se afle valoarea $F_{10,4;0,95}$.

Conform observaţiei 4 , avem

$$F_{10,4;0,95} = \frac{1}{F_{4,10;0,05}} .$$

In tabelul VIII din anexă, găsim $F_{4,10;0,05} = 3,478$. Prin urmare, $F_{10,4;0,95} = \frac{1}{3,478} \simeq 0,28$.

Teorema 14. Dacă variabilele aleatoare U şi V sînt independente şi urmează relartiţiile $\chi^2(m,\sigma)$, respectiv $\chi^2(n,\sigma)$, atunci variabila aleatoare

$$S = \frac{n}{m} \frac{U}{V}$$

urmează repartiţia Snedecor cu m şi n grade de libertate.

Demonstraţie. Fie

$$f_1(x) = \begin{cases} \dfrac{x^{\frac{m}{2}-1} e^{-\frac{x}{2\sigma^2}}}{2^{\frac{m}{2}} \sigma^m \Gamma\left(\frac{m}{2}\right)} & \text{dacă } x > 0 \\ 0 & \text{dacă } x \leqslant 0 \end{cases}$$

densitatea de repartiţie a lui U ,

$$f_2(y) = \begin{cases} \dfrac{y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}}{2^{\frac{n}{2}} \sigma^n \Gamma\left(\frac{n}{2}\right)} & \text{dacă } y > 0 \\ 0 & \text{dacă } y \leq 0 \end{cases}$$

densitatea de repartiţie a lui V şi $f(x,y)$ densitatea de repartiţie comună a variabilelor aleatoare U şi V. Din independenţa variabilelor aleatoare U şi V , rezultă

$$f(x,y)=f_1(x)f_2(y) = \frac{x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{x^2+y^2}{2\sigma^2}}}{2^{\frac{m+n}{2}} \sigma^{m+n} \Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

Fie transformarea

$$\tau : \begin{cases} x = \frac{m}{n} sv \\ y = v \end{cases}$$

al cărei iacobian este $J = \frac{m}{n} v$. Conform teoremei 1, densitatea de repartiţie comună a variabilelor aleatoare S şi V este

$$g(s,v) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} s^{\frac{m}{2}-1} v^{\frac{m+n}{2}-1} e^{-\frac{v}{2\sigma^2}\left(1+\frac{m}{n}s\right)}}{2^{\frac{m+n}{2}} \sigma^{m+n} \Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

Densitatea de repartiţie a lui S se obţine integrînd $g(s,v)$ în raport cu v. Avem

$$h(s)=\int_0^{+\infty} g(s,v)dv = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} s^{\frac{m}{2}-1}}{2^{\frac{m+n}{2}} \sigma^{m+n} \Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} v^{\frac{m+n}{2}-1} e^{-\frac{v}{2\sigma^2}\left(1+\frac{m}{n}s\right)} dv$$

In urma schimbării de variabilă

$$z = \frac{v}{2\sigma^2}\left(1 + \frac{m}{n} s\right)$$

se obţine

$$h(s) = \frac{(\frac{m}{n})^{\frac{m}{2}} \Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} s^{\frac{m}{2}-1} (1 + \frac{m}{n} s)^{-\frac{m+n}{2}}$$

pentru $s > 0$, ceea ce reprezintă densitatea repartiţiei Snedecor cu m şi n grade de libertate.

Observaţia 5. Folosind teoremele 10 şi 14 se poate demonstra că dacă variabilele aleatoare X_i, $1 \leq i \leq m$, şi Y_j, $1 \leq j \leq n$, sînt independente şi urmează aceeaşi repartiţie $N(0, \sigma^2)$, atunci variabila aleatoare

$$S = \frac{n}{m} \frac{X_1^2 + \ldots + X_m^2}{Y_1^2 + \ldots + Y_n^2}$$

urmează repartiţia Snedecor cu m şi n grade de libertate.

P R O B L E M E

1. Din producţia unui strung, a cărui probabilitate de rebut este de 0,04 se extrag, în vederea controlului calităţii, un număr de 1000 piese. Controlul se efectuează după schema bilei revenite. Fie X numărul de piese rebut ce se pot obţine. Se cere:

a) repartiţia probabilistică a lui X

b) probabilitatea evenimentului ca în cele 1000 piese extrase să să găsească cel puţin 7 piese rebut.

c) media şi dispersia lui X

d) funcţia caracteristică a lui X

2. Numărul apelurilor ce sosesc într-o centrală telefonică, într-un anumit interval de timp, este o variabilă aleatoare X, ce urmează o repartiţie Poisson. Stiind că, în timp ce o oră sosesc în centrală, în medie, 30 de apeluri, se cere :

a) media şi dispersia lui X

b) probabilitatea evenimentului ca în timp de un minut să sosească în centrală cel mult două apeluri.

c) funcţia caracteristică a lui X.

3. Se dă funcţia

$$F(x) = \begin{cases} 0 & x < 0 \\ k(1 - \cos 2x) & 0 \leqslant x < \frac{\pi}{2} \\ 1 & x \geqslant \frac{\pi}{2} \end{cases}$$

unde $k \in R$. Se cere :

a) să se determine k încît F(x) să fie funcţie de repartiţie pentru variabila aleatoare X .

b) densitatea de repartiţie a lui X

c) probabilitatea evenimentului ca X să ia valori în $\left[0, \frac{\pi}{4}\right]$.

4. Se dă funcţia

$$f(x) = \begin{cases} k(2 - |x|) & |x| < 2 \\ 0 & |x| \geqslant 2 \end{cases}$$

unde $k \in R$. Se cere:

a) să se determine k încît f(x) să fie densitate de repartiţie pentru variabila aleatoare X .

b) funcţia de repartiţie a lui X.

c) media şi dispersia lui X .

d) modul şi cvartilele lui X.

e) probabilitatea evenimentului ca X să ia valori în $[-1,1]$.

5. Probabilitatea de realizare a unui eveniment, în fiecare probă, este egală cu 0,03. Cîte probe trebuie efectuate pentru ca, cu o probabilitate cel mult egală cu 0,21, abaterea frecvenţei relative a evenimentului de la probabilitatea sa, să fie în valoare absolută, cel puţin egală cu 0,01 ?

6. Fie X şi Y variabile aleatoare independente avînd densităţile de repartiţie f(x), respectiv g(y). Să se arate că:

i) variabila aleatoare Z = X + Y are densitatea de repartiţie

$$h(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)dx = \int_{-\infty}^{+\infty} f(z-x)g(x)dx$$

ii) variabila aleatoare $U = \frac{X}{Y}$ are densitatea de repartiţie

$$r(u) = -\int_{-\infty}^{0} yf(uy)g(y)dy + \int_{0}^{+\infty} yf(uy)g(y)dy$$

iii) variabila aleatoare V = XY are densitatea de repartiţie

$$s(v) = -\int_{-\infty}^{0} \frac{1}{y} f\left(\frac{v}{y}\right)g(y)dy + \int_{0}^{+\infty} \frac{1}{y} f\left(\frac{v}{y}\right)g(y)dy$$

Capitolul II

REPARTIŢII STATISTICE

§ 1. COLECTIVITATE STATISTICA.CARACTERISTICI.

Ca ramură a matematicii aplicate, statistica matematică,se ocupă cu gruparea, analiza şi interpretarea datelor experimentale cu privire la un anumit fenomen aleator de masă, în scopul obţinerii unor previziuni asupra desfăşurării viitoare a acestuia. Fundamentîndu-şi metodele proprii de cercetare pe teoria probabilităţilor, statistica matematică studiază colectivităţile statistice.

Colectivitatea statistică este o mulţime de elemente ce posedă o trăsătură comună. Elementele ce constituie o colectivitate statistică se numesc unităţi statistice sau indivizi. Volumul unei colectivităţi statistice este numărul indivizilor ce o constituie. O colectivitate statistică poate fi cercetată după una sau mai multe caracteristici. Prin caracteristică a unei colectivităţi statistice se înţelege o anumită proprietate urmărită la indivizii ei, în procesul prelucrării statistice. Caracteristica ce poate fi măsurată se numeşte caracteristică centitativă sau variabilă. Caracteristica ce nu poate fi măsurată se numeşte caracteristică calitativă sau atribut. O caracteristică cantitativă poate fi discretă sau continuă după cum ea poate lua o mulţime cel mult numărabilă de valori sau orice valoare dintr-un anumit interval. In cele ce urmează, prin colectivitate vom înţelege o colectivitate statistică.

Cînd se studiază diferite caracteristici ale unei colectivităţi, în majoritatea cazurilor, nu se pot cerceta decît părţi

ale colectivităţii generale, cu ajutorul selecţiilor sau eşantioanelor. Prin selecţie sau eşantion se înţelege o colectivitate parţială de elemente, extrase la întîmplare din colectivitatea generală, în scopul cercetării lor din punctul de vedere al unei caracteristici. Numărul elementelor dintr-o selecţie poartă numerele de volumul selecţiei . O selecţie se numeşte repetată dacă individul extras este reintrodus în colectivitate înainte de extragerea următorului. O selecţie se numeşte nerepetată dacă individul extras nu se reintroduce în colectivitate înainte de extragerea următorului. Dacă volumul colectivităţii este mult mai mare decît volumul selecţiei, practic, deosebirea între cele două feluri de selecţie este foarte mică şi în aplicaţii o selecţie nerepetată poate fi considerată ca o selecţie repetată, ceea ce face ca selecţia nerepetată să prezinte interes numai atunci cînd volumul colectivităţii este finit.

§ 2. REPARTIŢII STATISTICE UNIDIMENSIONALE.

Presupunem că, în urma unei selecţii de volum n, efectuată dintr-o colectivitate, la indivizii căreia se urmăreşte caracteristica X, s-au obţinut rezultatele $x_1,\dots,x_i,\dots,x_k$, unde $k \leqslant n$.

Prin frecvenţă absolută a valorii x_i se înţelege numărul indivizilor din selecţie, care au ca valoare a caracteristicii X, valoarea x_i . Dacă n_i este frecvenţa absolută a valorii x_i, atunci $\sum_{i=1}^{k} n_i = n$.

Prin frecvenţă relativă a valorii x_i se înţelege raportul dintre frecvenţa absolută a acestei valori şi volumul selecţiei. Dacă f_i este frecvenţa relativă a valorii x_i, deci

$f_i = \frac{n_i}{n}$, atunci $\sum_{i=1}^{k} f_i = 1$.

Uneori, în locul frecvenţei relative f_i, se foloseşte numărul de procente corespunzător valorii x_i, definit prin $\theta_i = 100\, f_i$. Evident, $\sum_{i=1}^{k} \theta_i = 100$.

Mulţimea valorilor obţinute într-o selecţie împreună cu frecvenţele absolute sau relative corespunzătoare determină o repartiţie statistică a caracteristicii X după care se cercetează colectivitatea. Ţinînd seama de faptul că fiecare selecţie furnizează o repartiţie statistică, urmează că X are mai multe repartiţii statistice. Intrucît colectivitatea se cercetează după o singură caracteristică, repartiţiile statistice ale lui X, se numesc unidimensionale. O repartiţie statistică se dă, de obicei, printr-un tablou de repartiţie, format din două linii, în care se scriu, pe prima linie valorile obţinute în selecţie, iar pe a doua linie frecvenţele absolute sau relative corespunzătoare acestora. Pentru o caracteristică cantitativă se pot distinge trei tipuri de repartiţii statistice.

O selecţie de volum n, cu toate valorile distincte, determină o repartiţie statistică de tipul I. Tablourile de repartiţie a frecvenţelor absolute şi relative sînt respectiv de forma

$$\begin{bmatrix} x_i \\ 1 \end{bmatrix}, \quad \begin{bmatrix} x_i \\ \frac{1}{n} \end{bmatrix}, \quad 1 \leqslant i \leqslant n$$

O selecţie de volum n, cu numai k valori distincte, $(k < n)$, determină o repartiţie statistică de tipul II. Tablourile de repartiţie a frecvenţelor absolute şi relative sînt respectiv de forma

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}, \quad \begin{bmatrix} x_i \\ f_i \end{bmatrix}, \quad 1 \leqslant i \leqslant k$$

unde $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} f_i = 1$.

Dacă volumul n al selecţiei este foarte mare sau caracteristica cantitativă este continuă, atunci valorile de selecţie se grupează în k intervalele disjuncte, ($k < n$), de lungimi egale sau neegale. In acest caz, toate valorile de selecţie, ce aparţin intervalului de grupare $(\ell_{i-1}, \ell_i]$ se consideră egale cu x_i, unde x_i se ia egal cu mijlocul intervalului şi deci

$$x_i = \frac{\ell_{i-1} + \ell_i}{2} \quad , \quad 1 \leqslant i \leqslant k$$

Prin frecvenţă absolută a intervalului $(\ell_{i-1}, \ell_i]$ se înţelege numărul valorilor de selecţie care aparţin acestuia.

Prin frecventă relativă a intervalului $(\ell_{i-1}, \ell_i]$ se înţelege raportul dintre frecvenţa absolută a intervalului şi volumul selecţiei. O asemenea selecţie, cu valorile grupate pe intervale, determină o repartiţie statistică de tipul III. Tablourile de repartiţie a frecvenţelor absolute şi relative au aceeaşi formă ca în cazul unei repartiţii statistice de tipul II, cu specificarea că x_i este mijlocul intervalului $(\ell_{i-1}, \ell_i]$.

O repartiţie statistică de tipul I sau tipul II este reprezentată grafic de poligomul frecvenţelor absolute sau relative, care se obţine, unind punctele de coordonate (x_i, n_i) sau (x_i, f_i), $1 \leqslant i \leqslant k$, prin segmente de dreaptă.

O repartiţie statistică de tipul III, a unei caracteristici continue, este reprezentată grafic, fie de histograma frecvenţelor absolute sau relative, fie de poligonul frecvenţelor absolute sau relative.

Pentru a construi histograma frecvenţelor absolute, se consideră pe axa absciselor intervalele de grupare şi pe fiecare asemenea interval se construieşte un dreptunghi de arie egală sau proporţională cu frecvenţa absolută a intervalului. Analog se construieşte histograma frecvenţelor relative.

Pentru a construi poligonul frecvenţelor absolute, se consideră pe axa absciselor intervalele de grupare şi în mijlocul fiecărui interval se ridică un segment de perpendiculară egal sau proporţional, în mărime, cu frecvenţa absolută a intervalului. Unind extremităţile superioare ale segmentelor corespunzătoare tuturor intervalelor se obţine poligonul frecvenţelor absolute. Analog se construieşte poligonul frecvenţelor relative.

Din histograma frecvenţelor se poate obţine poligonul frecvenţelor şi reciproc. Reprezentările grafice ale repartiţiilor statistice sînt utile pentru că ele sugerează reprezentarea grafică a repartiţiei probabilistice a caracteristicii după care se cercetează colectivitatea. Determinarea repartiţiei probabilistice a caracteristicii pe baza unei repartiţii statistice poartă numele de ajustare a repartiţiei statistice.

§ 3. MEDII STATISTICE.

Vom defini pentru repartiţiile statistice de tipul I şi tipul II mai multe feluri de medii pe care le numim medii statistice.

Media aritmetică ($\overline{x}$)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \quad \text{cu} \quad \sum_{i=1}^{k} n_i = n$$

In cele ce urmează vom demonstra cîteva proprietăţi ale mediei artimetice a unei repartiţii statistice de tipul II. Evident, aceste proprietăţi au loc şi pentru media aritmetică a unei repartiţii statistice de tipul I .

Media aritmetică posedă următoarele proprietăţi :

1. Media aritmetică este cuprinsă între cea mai mică şi cea mai mare valoare a repartiţiei statistice.

A demonstra această proprietate, înseamnă a arătă că, dacă $x_{min.}$ şi $x_{max.}$ sînt cea mai mică, respectiv cea mai mare valoare a repartiţiei statistice, atunci

$$x_{min.} < \bar{x} < x_{max.}$$

Intr-adevăr, pe de o parte avem

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i > \frac{1}{n}\sum_{i=1}^{k} x_{min.} n_i = x_{min.}$$

şi pe de altă parte, avem

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i < \frac{1}{n}\sum_{i=1}^{k} x_{max.} n_i = x_{max.}$$

2. Media aritmetică este <u>translativă</u>.

Fie repartiţia statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

cu media aritmetică $\bar{x}$ şi „a" un număr real oarecare. Se consideră o nouă repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i + a \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

cu media aritmetică $\bar{x}'$. Proprietatea de demonstrat revine la a arăta că $\bar{x}' = \bar{x} + a$

Intr-adevăr, avem

$$\bar{x}' = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i + a) = \frac{1}{n}\sum_{i=1}^{k} x_i n_i + a = \bar{x} + a$$

3. Media aritmetică este o <u>funcţie omogenă de gradul I</u> în valorile de selecţie.

Fie repartiţia statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

cu media aritmetică $\bar{x}$ şi t un număr real oarecare. Se consideră o nouă repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} tx_i \\ n_i \end{bmatrix}_{1 \leqslant i \leqslant k}$$

cu media aritmetică $\bar{x}'$. Proprietatea de demonstrat revine la a arăta că

$$\bar{x}' = t\bar{x}$$

Intr-adevăr, avem

$$\bar{x}' = \frac{1}{n} \sum_{i=1}^{k} n_i tx_i = t \cdot \frac{1}{n} \sum_{i=1}^{k} n_i x_i = t\bar{x} .$$

4. Suma abaterilor valorilor de selecţie de la media aritmetică a repartiţiei statistice corespunzătoare este nulă.

Intr-adevăr, avem

$$\sum_{i=1}^{k} n_i (x_i - \bar{x}) = \sum_{i=1}^{k} n_i x_i - \bar{x} \sum_{i=1}^{k} n_i = n\bar{x} - \bar{x}n = 0$$

5. Suma pătratelor abaterilor valorilor de selecţie de la o valoare oarecare reală este minimă cînd acea valoare este egală cu media aritmetică a repartiţiei statistice corespunzătoare.

Fie „a" o valoare reală oarecare. Avem de arătat că:

$$\sum_{i=1}^{k} n_i (x_i - a)^2 \geqslant \sum_{i=1}^{k} n_i (x_i - \bar{x})^2$$

Intr-adevăr, avem

$$\sum_{i=1}^{k} n_i (x_i - a)^2 = \sum_{i=1}^{k} n_i \left[(x_i - \bar{x}) + (\bar{x} - a)\right]^2 =$$

$$= \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 + 2(\bar{x} - a) \sum_{i=1}^{k} n_i (x_i - \bar{x}) + n(\bar{x} - a)^2 .$$

Ţinînd seama de proprietatea 4, ultima relaţie devine

$$\sum_{i=1}^{k} n_i (x_i - a)^2 = \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 + n(\bar{x}-a)^2 \geqslant \sum_{i=1}^{k} n_i (x_i - \bar{x})^2$$

Scanned with OKEN Scanner

Media armonică ($\bar{x}_h$)

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \quad , \quad \bar{x}_h = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}} \text{ cu } \sum_{i=1}^{k} n_i = n .$$

Media geometrică ($\bar{x}_g$)

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \dots x_n} \quad , \quad \bar{x}_g = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}} \text{ cu } \sum_{i=1}^{k} n_i = n .$$

Media de ordin r, ($r \in R$)

$$\bar{x}_{(r)} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^r \right)^{\frac{1}{r}} \quad , \quad \bar{x}_{(r)} = \left(\frac{1}{n} \sum_{i=1}^{k} n_i x_i^r \right)^{\frac{1}{r}} \quad , \quad \sum_{i=1}^{k} n_i = n$$

Observaţia 1. Dacă $x_i > 0$, $1 \leq i \leq k$, atunci

$$\bar{x}_g = \lim_{r \to 0} \bar{x}_{(r)} .$$

Dacă notăm $G = \lim_{r \to 0} x_{(r)}$, avem de arătat că $G = \bar{x}_g$.

Ţinînd seama de definiţia lui $\bar{x}_{(r)}$, notaţia făcută devine

$$G = \lim_{r \to 0} \left(\frac{1}{n} \sum_{i=1}^{k} x_i^r n_i \right)^{\frac{1}{r}} .$$

Avem

$$\ln G = \lim_{r \to 0} \frac{\ln \left(\frac{1}{r} \sum_{i=1}^{k} x_i^r n_i \right)}{r} = \lim_{r \to 0} \frac{\sum_{i=1}^{k} n_i x_i^r \ln x_i}{\sum_{i=1}^{k} x_i^r n_i} =$$

$$= \frac{\sum_{i=1}^{k} n_i \ln x_i}{n} = \ln \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}}$$

de unde

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}}$$

adică

$$G = \bar{x}_g$$

Observaţia 2. Se poate demonstra (a se vedea Ionescu H.M. [6]) că dacă $x_i > 0$, $1 \leq i \leq k$, atunci oricare ar fi numerele reale r,s cu $r \leq s$, are loc

$$\bar{x}_{(r)} \leq \bar{x}_{(s)}$$

Particularizînd pe r în definiţia lui $\bar{x}_{(r)}$ şi ţinînd seama de observaţia 1, avem

$$\bar{x}_{(-1)} = \bar{x}_h \ , \ \bar{x}_{(0)} = \bar{x}_g \ , \ \bar{x}_{(1)} = \bar{x}$$

Ţinînd seama de observaţia 2, rezultă că, între mediile armonică, geometrică şi aritmetică, au loc inegalităţile

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x}$$

§ 4. MOMENTE STATISTICE .

Prin momente statistice vom înţelege momentele unei repartiţii statistice. Fie „a" un număr real oarecare.

Momentul de ordin h, (h = 0,1,2,...), în raport cu valoarea „a" , se defineşte, pentru o repartiţie statistică de tipul I, prin

$$\bar{M}_h^{(a)} = \frac{1}{n} \sum_{i=1}^{n} (x_i - a)^h$$

iar pentru o repartiţie statistică de tipul II, prin

$$\bar{M}_h^{(a)} = \frac{1}{n} \sum_{i=1}^{k} (x_i - a)^h n_i$$

Dacă a = 0, vom nota

$$\bar{M}_h^{(o)} = \bar{\alpha}_h$$

unde

$$\bar{\alpha}_h = \frac{1}{n} \sum_{i=1}^{n} x_i^h$$

respectiv

$$\bar{\alpha}_h = \frac{1}{n} \sum_{i=1}^{k} x_i^h n_i$$

este numit moment iniţial de ordin h . Evident, $\bar{\alpha}_o = 1$, $\bar{\alpha}_1 = \bar{x}$.

Dacă $a = \bar{x}$, vom nota $\bar{M}_h^{(\bar{x})} = \bar{\mu}_h$

unde

$$\bar{\mu}_h = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^h$$

respectiv

$$\bar{\mu}_h = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^h n_i$$

Scanned with OKEN Scanner

este numit <u>moment centrat de ordin h</u> . Evident, $\bar{\mu}_0 = 1$, $\bar{\mu}_1 = 0$.

Pentru o repartiţie statistică de tipul III, formulele ce definesc momentele, sînt aceleaşi ca pentru o repartiţie statistică de tipul II, cu specificarea că valoarea x_i este mijlocul intervalului $(\ell_{i-1}, \ell_i]$.

<u>Observaţia 1</u>. Momentele centrate sînt invariante la o translaţie a valorilor repartiţiei statistice.

Vom demonstra acest lucru pentru o repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

pentru care , $\bar{\mu}_h$ este momentul centrat de ordin h. Se consideră o nouă repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} y_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

unde

$$y_i = x_i - a \quad , \quad 1 \leq i \leq k$$

iar „a" este un număr real orrecare. Fie $\bar{\mu}'_h$ momentul centrat al repartiţiei statistice cu valorile y_i, adică,

$$\bar{\mu}'_h = \frac{1}{n} \sum_{i=1}^{k} (y_i - y)^h n_i \ .$$

Vom arăta că, $\bar{\mu}'_h = \bar{\mu}_h$ pentru $h = 0,1,2,\ldots$. Intr-adevăr, folosind faptul că media aritmetică este translativă, avem

$$\bar{y} = \bar{x} - a$$

Atunci

$$\bar{\mu}'_h = \frac{1}{n} \sum_{i=1}^{k} (x_i - a - \bar{x} + a)^h n_i = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^h n_i = \bar{\mu}_h$$

pentru $h = 0,1,2,\ldots$

Intre momentele statistice iniţiale şi centrate există

aceleaşi formule de legătură ca şi între momentele iniţiale şi centrate teoretice date în § 1, cap.I.

Dacă se cunosc momentele iniţiale de ordin r , (r = 0,1,2,..h), atunci momentul centrat de ordin h, se exprimă prin

$$(1) \qquad \bar{\mu}_h = \sum_{j=0}^{h} (-1)^j C_h^j \bar{\alpha}_{h-j} \bar{\alpha}_1^j .$$

Vom demonstra formula (1) în cazul unei repartiţii statistice de tipul II. Intr-adevăr, avem

$$\bar{\mu}_h = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^h n_i = \frac{1}{n} \sum_{i=1}^{k} n_i \left[\sum_{j=0}^{h} (-1)^j C_h^j x_i^{h-j} \bar{x}^j \right] =$$

$$= \sum_{j=0}^{h} (-1)^j C_h^j \left[\frac{1}{n} \sum_{i=0}^{k} x_i^{h-j} n_i \right] \bar{x}^j = \sum_{j=0}^{h} (-1)^j C_h^j \bar{\alpha}_{h-j} \bar{\alpha}_1^j$$

Pentru h = 0,1,2,3,4, din formula (1) se obţine :

$$\bar{\mu}_0 = 1, \bar{\mu}_1 = 0, \bar{\mu}_2 = \bar{\alpha}_2 - \bar{\alpha}_1^2 , \bar{\mu}_3 = \bar{\alpha}_3 - 3\bar{\alpha}_2\bar{\alpha}_1 + 2\bar{\alpha}_1^3$$

$$\bar{\mu}_4 = \bar{\alpha}_4 - 4\bar{\alpha}_3\bar{\alpha}_1 + 6\bar{\alpha}_2\bar{\alpha}_1^2 - 3\bar{\alpha}_1^4 .$$

Dacă se cunosc $\bar{\alpha}_1$ şi momentele centrate de ordin r, (r = 0,1,2,...,h), atunci momentul iniţial de ordin h, se exprimă prin

$$(2) \qquad \bar{\alpha}_h = \sum_{j=0}^{h} C_h^j \bar{\mu}_{h-j} \bar{\alpha}_1^j$$

Vom demonstra formula (2) în cazul unei repartiţii statistice de tipul II. Intr-adevăr, avem

$$\bar{\alpha}_h = \frac{1}{n} \sum_{i=1}^{k} x_i^h n_i = \frac{1}{n} \sum_{i=1}^{k} \left[(x_i - \bar{x}) + \bar{x} \right]^h n_i = \frac{1}{n} \sum_{i=1}^{k} n_i \left[\sum_{j=0}^{h} C_h^j (x_i - \bar{x})^{h-j} \bar{x}^j \right] =$$

$$= \sum_{j=0}^{h} C_h^j \left[\frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^{h-j} n_i \right] \bar{x}^j = \sum_{j=0}^{h} C_h^j \bar{\mu}_{h-j} \bar{\alpha}_1^j$$

Metoda zeroului fals pentru calculul momentelor unei repartiţii statistice de tipul II.

Fie o repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

cu $\sum_{i=1}^{k} n_i = n$ şi „a" una din valorile repartiţiei, numită zerou fals.

Se consideră o nouă repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} y_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

unde

$$y_i = x_i - a , \quad 1 \leq i \leq k .$$

Dacă se notează momentele iniţiale ale celor două repartiţii statistice cu $\bar{\alpha}_h$, respectiv $\bar{\alpha}'_h$, atunci are loc

$$\bar{\alpha}_h = \sum_{j=0}^{h} C_h^j \bar{\alpha}'_{h-j} a^j . \tag{3}$$

Intr-adevăr,

$$\bar{\alpha}_h = \frac{1}{n} \sum_{i=1}^{k} x_i^h n_i = \frac{1}{n} \sum_{i=1}^{k} (y_i+a)^h n_i = \frac{1}{n} \sum_{i=1}^{k} n_i \left(\sum_{j=0}^{h} C_h^j y_i^{h-j} a^j \right) =$$

$$= \sum_{j=0}^{h} C_h^j \left(\frac{1}{n} \sum_{i=1}^{k} y_i^{h-j} n_i \right) a^j = \sum_{j=0}^{h} C_h^j \bar{\alpha}'_{h-j} a^j .$$

Pentru $h = 0,1,2,3,4$, din formula (3) se obţine :

$$\bar{\alpha}_0 = 1, \quad \bar{\alpha}_1 = \bar{\alpha}'_1 + a , \quad \bar{\alpha}_2 = \bar{\alpha}'_2 + 2\bar{\alpha}'_1 a + a^2$$

$$\bar{\alpha}_3 = \bar{\alpha}'_3 + 3\bar{\alpha}'_2 a + 3\bar{\alpha}'_1 a^2 + a^3, \quad \bar{\alpha}_4 = \bar{\alpha}'_4 + 4\bar{\alpha}'_3 a + 6\bar{\alpha}'_2 a^2 + 4\bar{\alpha}'_1 a^3 + a^4 .$$

Zeroul fals se alege în aşa fel încît valorile y_i, $1 \leq i \leq k$, să fie cît mai mici. De obicei, zeroul fals „a" se alege apropiat de $\bar{x}$.

Dacă se notează momentele centrate ale celor două repartiţii statistice cu $\bar{\mu}_h$, respectiv $\bar{\mu}'_h$, atunci are loc

$$\bar{\mu}_h = \bar{\mu}'_h \tag{4}$$

pentru că am demonstrat că momentele centrate sînt invariante la o translaţie a valorilor repartiţiei statistice. Din relaţiile (1) şi (4) rezultă

$$(5) \qquad \bar{\mu}_h = \sum_{j=0}^{h} (-1)^j C_h^j \bar{\alpha}'_{h-j} \bar{\alpha}'^j_1$$

Pentru $h = 0,1,2,3,4$ din formula (5) se obţine :

$$\bar{\mu}_0 = 1 \,, \quad \bar{\mu}_1 = 0, \quad \bar{\mu}_2 = \bar{\alpha}'_2 - \bar{\alpha}'^2_1, \quad \bar{\mu}_3 = \bar{\alpha}'_3 - 3\bar{\alpha}'_2\bar{\alpha}'_1 + 2\bar{\alpha}'^2_1$$

$$\bar{\mu}_4 = \bar{\alpha}'_4 - 4\bar{\alpha}'_3\bar{\alpha}_1 + 6\bar{\alpha}'_2\bar{\alpha}'^2_1 - 3\bar{\alpha}'^4_1$$

Metoda zeroului fals pentru calculul momentelor unei repartiţii statistice de tipul III.

Fie o repartiţie statistică de tipul III, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leqslant i \leqslant k}$$

cu $\sum_{i=1}^{k} n_i = n$ şi $x_i = \frac{\ell_{i-1} + \ell_i}{2}$, $1 \leqslant i \leqslant k$, iar intervalele de grupare au lungimea comună egală cu d . Intrucît x_i este mijlocul intervalului $(\ell_{i-1}, \ell_i]$ şi toate cele k intervale au aceeaşi lungime egală cu d, rezultă că ,

$$x_i - x_{i-1} = d \quad , \quad 2 \leqslant i \leqslant k$$

de unde

$$x_i = x_{i-1} + d \,, \quad 2 \leqslant i \leqslant k$$

Evident

$$x_i = x_1 + (i-1)d, \quad 1 \leqslant i \leqslant k$$

Alegem zeroul fals „a" „ egal cu una din valorile repartiţiei. ceea ce înseamnă că

$$a = x_1 + rd$$

unde r este un număr întreg.

Se consideră o nouă repartiţie statistică de tipul III, dată de tabloul de repartiţie

$$\begin{bmatrix} y_i \\ n_i \end{bmatrix}_{1 \leqslant i \leqslant k}$$

unde

$$y_i = \frac{x_i - a}{d} \quad , \quad 1 \leqslant i \leqslant k \,.$$

Din relaţiile

$$x_i = x_1 + (i-1)d$$

$$a = x_1 + rd$$

rezultă

$$x_i - a = (i-1) - r\,d$$

deci

$$\frac{x_i - a}{d} = y_i = i - 1 - r\,, \quad 1 \leqslant i \leqslant k$$

unde $i - 1 - r$ este un număr întreg. Prin urmare, valorile y_i ale noii repartiţii statistice sînt numere întregi.

Notăm momentele iniţiale ale celor două repartiţii statistice cu $\bar{\alpha}_h$, respectiv $\hat{\alpha}_h$ şi $b = \frac{a}{d}$. Cu aceste notaţii, are loc

(6) $$\bar{\alpha}_h = d^h\left(\sum_{j=0}^{h} C_h^j \hat{\alpha}_{h-j} b^j\right)$$

Intr-adevăr,

$$\bar{\alpha}_h = \frac{1}{n}\sum_{i=1}^{k} x_i^h n_i = \frac{d^h}{n}\sum_{i=1}^{k} n_i\left[\frac{x_i - a}{d} + \frac{a}{d}\right]^h = \frac{d^h}{n}\sum_{i=1}^{k} n_i (y_i + b)^h =$$

$$= \frac{d^h}{n}\sum_{i=1}^{k} n_i\left[\sum_{j=0}^{h} C_h^j y_i^{h-j} b^j\right] = d^h \sum_{j=0}^{h} C_h^j \left(\frac{1}{n}\sum_{i-1}^{k} y_i^{h-j} n_i\right) b^j =$$

$$= d^h \sum_{j=0}^{h} C_h^j \hat{\alpha}_{h-j} b^j .$$

Pentru $h = 0,1,2,3,4$ din formula (6) se obţine :

$$\bar{\alpha}_0 = 1, \bar{\alpha}_1 = d(\hat{\alpha}_1 + b), \bar{\alpha}_2 = d^2(\hat{\alpha}_2 + 2\hat{\alpha}_1 b + b^2)$$

$$\bar{\alpha}_3 = d^3(\hat{\alpha}_3 + 3\hat{\alpha}_2 b + 3\hat{\alpha}_1 b^2 + b^3)$$

$$\bar{\alpha}_4 = d^4(\hat{\alpha}_4 + 4\hat{\alpha}_3 b + 6\hat{\alpha}_2 b^2 + 4\hat{\alpha}_1 b^3 + b^4)$$

Vom demonstra că, are loc

(7) $$\bar{\mu}_h = d^h \sum_{j=0}^{h} (-1)^j C_h^j \hat{\alpha}_{h-j} \hat{\alpha}_1^j$$

Din formula (6), pentru $h = 1$, s-a obţinut

$$\frac{\bar{\alpha}_1}{d} = \frac{\bar{x}}{d} = \hat{\alpha}_1 + b$$

Ţinînd seama de ultima relaţie, avem

$$\bar{\mu}_h = \frac{1}{n}\sum_{i=1}^{k}(x_i-\bar{x})^h n_i = \frac{d^h}{n}\sum_{i=1}^{k}\left(\frac{x_i-a}{d} - \frac{\bar{x}-a}{d}\right)^h n_i =$$

$$= \frac{d^h}{n}\sum_{i=1}^{k}\left(y_i - \frac{\bar{x}}{d} + \frac{a}{d}\right)^h n_i = \frac{d^h}{n}\sum_{i=1}^{k}\left(y_i - \frac{\bar{x}}{d} + b\right)^h n_i =$$

$$= \frac{d^h}{n}\sum_{i=1}^{k}(y_i - \hat{\alpha}_1 - b + b)^h n_i = \frac{d^h}{n}\sum_{i=1}^{k}(y_i - \hat{\alpha}_1)^h n_i =$$

$$= \frac{d^h}{n}\sum_{i=1}^{k} n_i \left[\sum_{j=0}^{h}(-1)^j C_h^j y_i^{h-j}\hat{\alpha}_1^j = d^h \sum_{j=0}^{h}(-1)^j C_h^j \left(\frac{1}{n}\sum_{i=1}^{k} y_i^{h-j} n_i\right)\hat{\alpha}_1^j =\right.$$

$$= d^h \sum_{j=0}^{h}(-1)^j C_h^j \hat{\alpha}_{h-j}\hat{\alpha}_1^j \ .$$

Pentru $h = 0,1,2,3,4$ din formula (7) se obţine :

$$\bar{\mu}_0 = 1 \ , \quad \bar{\mu}_1 = 0 \ , \quad \bar{\mu}_2 = d^2(\hat{\alpha}_2 - \hat{\alpha}_1^2)$$

$$\bar{\mu}_3 = d^3\left(\hat{\alpha}_3 - 3\hat{\alpha}_2\hat{\alpha}_1 + 2\hat{\alpha}_1^3\right)$$

$$\bar{\mu}_4 = d^4\left(\hat{\alpha}_4 - 4\hat{\alpha}_3\hat{\alpha}_1 + 6\hat{\alpha}_2\hat{\alpha}_1^2 - 3\hat{\alpha}_1^4\right).$$

Observaţia 2. Considerarea tuturor valorilor din intervalul $(\ell_{i-1}, \ell_i]$ egale cu mijlocul acestui interval, pentru o repartiţie statistică de tipul III, conduce la erori. Din acest motiv, momentelor iniţiale şi centrate li se aduc corecţii, cunoscute sub numele de „corecţiile lui Sheppard". Dacă se notează momentele iniţiale şi centrate corectate cu $\tilde{\alpha}_h$, respectiv $\tilde{\mu}_h$, atunci au loc relaţiile :

$$\tilde{\alpha}_1 = \bar{\alpha}_1 \qquad \tilde{\mu}_1 = \bar{\mu}_1$$

$$\tilde{\alpha}_2 = \bar{\alpha}_2 + \frac{d^2}{12} \qquad \tilde{\mu}_2 = \bar{\mu}_2 - \frac{d^2}{12}$$

$$\tilde{\alpha}_3 = \bar{\alpha}_3 + \frac{d^2}{4}\bar{\alpha}_1 \qquad \tilde{\mu}_3 = \bar{\mu}_3$$

$$\tilde{\alpha}_4 = \bar{\alpha}_4 + \frac{d^2}{2}\bar{\alpha}_2 + \frac{d^4}{80} \qquad \tilde{\mu}_4 = \bar{\mu}_4 - \frac{d^2}{2}\bar{\mu}_2 + \frac{7d^4}{240}$$

Scanned with OKEN Scanner

§ 5. FUNCŢIA DE FRECVENŢĂ CUMULATĂ.

Dintr-o colectivitate, la indivizii căreia se urmăreşte caracteristica X, se consideră o selecţie de volum n . Fie x o valoare reală fixată .

Prin frecvenţă absolută (relativă) cumulată crescătoare a valorii x se înţelege suma frecvenţelor absolute (relative) ale tuturor valorilor, din selecţia considerată, mai mici decît x .

Frecvenţa relativă cumulată crescătoare a unei valori, mai poate fi definită ca, raportul dintre frecvenţa absolută cumulată crescătoare a acelei valori şi volumul selecţiei.

Prin frecvenţă absolută (relativă) cumulată descrescătoare a valorii x se înţelege suma frecvenţelor absolute (relative) ale tuturor valorilor, din selecţia considerată, mai mari sau cel puţin egale cu x .

In cele ce urmează, prin frecvenţă cumulată a unei valori, vom înţelege frecvenţa relativă cumulată crescătoare a acelei valori.

Prin funcţie de frecvenţă cumulată se înţelege o aplicaţie

$$\overline{F}_n : R \longrightarrow R$$

care pentru x fixat este egală cu frecvenţa relativă cumulată crescătoare a valorii x, într-o selecţie de volum n. Dacă pentru x fixat, se notează cu n_x, frecvenţa absolută cumulată crescătoare a valorii x , atunci funcţia de frecvenţă cumulată este

$$\overline{F}_n(x) = \frac{n_x}{n}$$

Pentru o repartiţie statistică de tipul I, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ 1 \end{bmatrix}_{1 \leq i \leq n}$$

în presupunerea $x_1 < x_2 < \ldots < x_i < \ldots < x_n$, funcţia de frecvenţă cumulată este

$$\bar{F}_n(x) = \begin{cases} 0 & x \leq x_1 \\ \frac{i-1}{n} & x_{i-1} < x \leq x_i, \ (i=2,3,\ldots,n) \\ 1 & x > x_n \end{cases}$$

Intr-adevăr, dacă $x \leq x_1$, atunci $n_x = 0$ şi deci $\bar{F}_n(x) = \frac{n_x}{n} = 0$ Dacă $x_{i-1} < x \leq x_i$, $(i = 2,3,\ldots,n)$, atunci $n_x = i - 1$ şi deci $\bar{F}_n(x) = \frac{i-1}{n}$. Dacă $x > x_n$, atunci $n_x = n$ şi deci $\bar{F}_n(x) = 1$. In oricare punct $x_i (i = 1,2,\ldots,n)$ $\bar{F}_n(x)$ este continuă numai la stînga, avînd saltul egal cu $\frac{1}{n}$. Graficul lui $\bar{F}_n(x)$ este o curbă în trepte. Toate treptele au înălţimea $\frac{1}{n}$.

Pentru o repartiţie statistică de tipul II, dată de tabloul de repartiţie

$$\begin{bmatrix} x_i \\ n_i \end{bmatrix}_{1 \leq i \leq k}$$

în presupunerea $x_1 < x_2 < \ldots < x_i < \ldots < x_k$, funcţia de frecvenţă cumulată este

$$\bar{F}_n(x) = \begin{cases} 0 & x \leq x_1 \\ \sum_{j=1}^{i-1} f_j & x_{i-1} < x \leq x_i, \ (i=2,\ldots,k) \\ 1 & x > x_k \end{cases}$$

Intr-adevăr, dacă $x \leq x_1$, atunci $n_x = 0$ şi deci $\bar{F}_n(x) = 0$. Dacă $x_{i-1} < x \leq x_i$, $(i = 2,3,\ldots,k)$, atunci $n_x = n_1 + n_2 + \ldots + n_{i-1} = \sum_{j=1}^{i-1} n_j$ şi deci $\bar{F}_n(x) = \sum_{j=1}^{i-1} \frac{n_j}{n} = \sum_{j=1}^{i-1} f_i$. Dacă $x > x_k$, atunci $n_x = n_1 + n_2 + \ldots + n_k = n$ şi deci $\bar{F}_n(x) = 1$. In oricare punct x_i, $(i = 1,2,\ldots,k)$, $\bar{F}_n(x)$ este continuă numai la stînga, avînd saltul egal cu f_i. Graficul lui $\bar{F}_n(x)$ este o curbă în trepte. Treptele au înălţimi diferite.

Pentru o repartiţie statistică de tipul III, cu intervalele

de grupare $(\ell_{i-1}, \ell_i]$, $(i = 1,2,\ldots,k)$, avînd lungimile egale cu d, funcţia de frecvenţă cumulată este

$$\bar{F}_n(x) = \begin{cases} 0 & x \leqslant \ell_0 \\ \sum_{j=1}^{i-1} f_j + \dfrac{x - \ell_{i-1}}{d} f_i & \ell_{i-1} < x \leqslant \ell_i,\ (i=1,..,k) \\ 1 & x > \ell_k \end{cases}$$

Intr-adevăr, dacă $x \leqslant \ell_0$, atunci $n_x = 0$ şi deci $\bar{F}_n(x)=0$. Dacă $x > \ell_k$, atunci $n_x = n$ şi deci $\bar{F}_n(x) = 1$. Dacă $\ell_{i-1} < x \leqslant \ell_i$, $(i = 1,2,\ldots,k)$ şi presupunem că între punctele $M_{i-1}(\ell_{i-1}, \sum_{j=1}^{i-1} f_j)$ şi $M_i(\ell_i, \sum_{j=1}^{i} f_j)$, funcţia de frecvenţă cumulată este liniară, scriind ecuaţia dreptei ce trece prin aceste puncte, se obţine

$$\frac{\bar{F}_n(x) - \sum_{j=1}^{i-1} f_j}{\sum_{j=1}^{i} f_j - \sum_{j=1}^{i-1} f_j} = \frac{x - \ell_{i-1}}{\ell_i - \ell_{i-1}} \quad \text{sau}$$

$$\frac{\bar{F}_n(x) - \sum_{j=1}^{i-1} f_j}{f_i} = \frac{x - \ell_{i-1}}{d}$$

de unde

$$\bar{F}_n(x) = \sum_{j=1}^{i-1} f_j + \frac{x - \ell_{i-1}}{d} f_i \ .$$

Graficul lui $\bar{F}_n(x)$ este o linie poligonală obţinută unind prin segmente de dreaptă punctele $M_{i-1}(\ell_{i-1}, \sum_{j=1}^{i-1} f_j)$, $(i = 1,2,\ldots,k+1)$.

Observaţie. Funcţia de frecvenţă cumulată are toate proprietăţile unei funcţii de repartiţie. Din acest motiv, funcţia de frecvenţă cumulată se mai numeşte şi funcţia de repartiţie statistică.

Pînă aici am văzut cum se determină funcţia de frecvenţă cumulată cînd se cunoaşte tabloul repartiţiei frecvenţelor. Reciproc, dacă se cunoaşte funcţia de frecvenţă cumulată se poate determina tabloul repartiţiei frecvenţelor.

Intr-adevăr, în cazul unei repartiţii statistice de tipul I sau tipul II valorile x_i sînt puncte de discontinuitate la dreapta ale lui $\bar{F}_n(x)$, iar frecvenţele relative f_i sînt salturile lui $\bar{F}_n(x)$ în punctele x_i. In cazul unei repartiţii statistice de tipul III, frecvenţa relativă a intervalului $(\ell_{i-1}, \ell_i]$ este dată de

$$f_i = \bar{F}_n(\ell_i) - \bar{F}_n(\ell_{i-1})$$

Intr-adevăr, ţinînd seama de expresia lui $\bar{F}_n(x)$ pentru o repartiţie statistică de tipul III, se obţine

$$\bar{F}_n(\ell_i) - \bar{F}_n(\ell_{i-1}) = \sum_{j=1}^{i-1} f_j + \frac{\ell_i - \ell_{i-1}}{d} f_i - \sum_{j=1}^{i-2} f_j - \frac{\ell_{i-1} - \ell_{i-2}}{d} f_{i-1} =$$

$$= f_{i-1} + f_i - f_{i-1} = f_i .$$

Prin urmare, o repartiţie statistică unidimensională, poate fi dată, fie prin tabloul de repartiţie a frecvenţelor absolute (relative), fie prin funcţia de frecvenţă cumulată.

§ 6. CVANTILE ŞI MODURI STATISTICE

Fie $\bar{F}_n(x)$ funcţia de frecvenţă cumulată şi f o frecvenţă relativă, deci, $0 < f < 1$.

Prin <u>f-cvantilă statistică</u> se înţelege orice valoare reală x_f pentru care au loc

$$\bar{F}_n(x_f) \leqslant f , \quad \bar{F}_n(x_f + 0) \geqslant f$$

Pentru un f dat, pot exista mai multe f-cvantile statistice. Dacă $\bar{F}_n(x)$ este continuă şi crescătoare, atunci pentru un f dat, există o singură f-cvantilă statistică, ce se obţine rezolvînd ecuaţia

$$\bar{F}_n(x) = f$$

Pentru $f = \frac{i}{4}$, $(i = 1,2,3)$, se obţin <u>cvartilele statistice</u> numite <u>cvartila inferioară</u>, <u>cvartila mijlocie (mediana)</u> şi <u>cvartila superioară</u>, notate respectiv $\bar{Q}_1, \bar{Q}_2$ sau $\bar{M}_e$ şi $\bar{Q}_3$.

Pentru $f = \frac{r}{10}$, $(r = 1,2,\ldots,9)$, se obţin <u>decilele statistice</u> notate $\bar{c}_r$, $(r = 1,2,\ldots,9)$.

Pentru $f = \frac{k}{100}$, $(k = 1,2,\ldots,99)$, se obţin <u>centilele statistice</u> notate $\bar{D}_k$, $(k = 1,2,\ldots,99)$.

Cea mai utilizată cvartilă statistică este mediana. Ţinînd seama de definiţia f-cvantilei stiatistice, mediana statistică satisface relaţiile

$$\bar{F}_n(\bar{M}_e) \leq \frac{1}{2}, \quad \bar{F}_n(\bar{M}_e + 0) \geq \frac{1}{2}.$$

Pentru o repartiţie statistică de tipul I, valoarea $\frac{1}{2}$ a lui $\bar{F}_n(x)$ este, în general, cuprinsă între $\bar{F}_n(x_i)$ şi $\bar{F}_n(x_{i+1})$, ceea ce revine la

$$\frac{i-1}{n} < \frac{1}{2} < \frac{i}{n}$$

sau

$$i - 1 < \frac{n}{2} < i$$

Dacă n este impar, fie $n = 2r + 1$, atunci $\frac{n}{2} = r + \frac{1}{2}$ nu este întreg şi deci $\left[\frac{n}{2}\right] = r$. Pe de altă parte, din

$$i - 1 < \frac{n}{2} < i$$

urmează că, $\left[\frac{n}{2}\right] = i - 1$, de unde $i = \left[\frac{n}{2}\right] + 1 = r + 1 = \frac{n+1}{2}$.

Prin urmare, dacă n este impar, mediana statistică este egală cu valoarea de rang $\frac{n+1}{2}$, unde am presupus că valorile repartiţiei statistice sînt ordonate crescător.

Dacă n este par, fie $n = 2r$, atunci $\frac{n}{2} = r$ este întreg şi relaţia

$$i - 1 < \frac{n}{2} < i$$

nu poate avea loc, pentru că între întregii consecutivi i-1 şi i nu poate exista alt întreg. In acest caz, de obicei, ca mediană statistică, se consideră semisuma valorilor de rang

$\frac{n}{2}$ şi $\frac{n}{2}+1$.

Pentru o repartiţie statistică de tipul II, faptul că valoarea $\frac{1}{2}$ a lui $\bar{F}_n(x)$ este cuprinsă între $\bar{F}_n(x_i)$ şi $\bar{F}_n(x_{i+1})$, revine la

$$\sum_{j=1}^{i-1} f_j < \frac{1}{2} < \sum_{j=1}^{i} f_j .$$

In acest caz, $\bar{M}_e = x_i$, unde i este cel mai mic rang pentru care are loc

$$\sum_{j=1}^{i} f_j > \frac{1}{2} .$$

Prin <u>interval median</u>, în cazul unei repartiţii statistice de tipul III, se înţelege acel interval în care se află valoarea de rang $\frac{n}{2}$, indiferent de paritatea lui n. Mediana statistică se află prin rezolvarea ecuaţiei

$$\bar{F}_n(x) = \frac{1}{2}$$

şi se obţine

$$\sum_{j=1}^{i-1} f_j + \frac{x-\ell_{i-1}}{d} f_i = \frac{1}{2} \quad \text{sau} \quad x = \ell_{i-1} + \frac{\frac{1}{2} - \sum_{j=1}^{i-1} f_j}{f_i} d$$

sau

$$x = \ell_{i-1} + \frac{n-2\sum_{j=1}^{i-1} n_j}{2n_i} d$$

unde

ℓ_{i-1} - limita inferioară a intervalului median

d - lungimea intervalului median

n_i - frecvenţa absolută a intervalului median

Cu ajutorul cvartilelor statistice se definesc <u>variaţia intercvartilică statistică</u> prin

$$\bar{Q} = \frac{\bar{Q}_3 - \bar{Q}_1}{2}$$

şi <u>coeficientul de variaţie intercvartilic statistic</u> prin

$$\bar{q} = \frac{\bar{Q}}{\bar{M}_e}$$

Prin mod statistic, pentru o repartiţie statistică de tipul I sau tipul II, se înţelege acea valoare a repartiţiei care are frecvenţa absolută (relativă) cea mai mare, dintre frecvenţele tuturor valorilor repartiţiei. Modul statistic îl notăm cu $\bar{M}_o$.

Prin interval modal, pentru o repartiţie statistică de tipul III, se înţelege intervalul care are frecvenţa absolută (relativă) cea mai mare dintre frecvenţele tuturor intervalelor repartiţiei.

Notăm

$\underline{\ell}$ - limita inferioară a intervalului modal

$\bar{\ell}$ - limita superioară a intervalului modal

Δ_1 - diferenţa dintre frecvenţa absolută (relativă) a intervalului modal şi frecvenţa absolută (relativă) a intervalului anterior celui modal

Δ_2 - diferenţa dintre frecvenţa absolută (relativă) a intervalului modal şi frecvenţa absolută (relativă) a intervalului următor celui modal

d - lungimea intervalului modal

Dacă frecvenţa absolută (relativă) a intervalului anterior celui modal este mai mică decît frecvenţa absolută (relativă) a intervalului următor celui modal atunci modul statistic $\bar{M}_o$ se calculează prin formula

$$\bar{M}_o = \bar{\ell} - \frac{\Delta_2}{\Delta_1 + \Delta_2} d$$

Dacă frecvenţa absolută (relativă) a intervalului anterior celui modal este mai mare decît frecvenţa absolută (relativă) a intervalului următor celui modal atunci modul statistic $\bar{M}_o$ se calculează prin formula

$$\bar{M}_o = \underline{\ell} + \frac{\Delta_1}{\Delta_1 + \Delta_2} d$$

Deoarece modul statistic ($\bar{M}_0$) , mediana statistică ($\bar{M}_e$) şi media aritmetică ($\bar{x}$) caracterizează tendinţa centrală a repartiţiei statistice, aceşti indicatori se numesc indicatori de poziţie.

§ 7. INDICATORII VARIAŢIEI. ASIMETRIE. EXCES.

Indicatorii variaţiei caracterizează împrăştierea valorilor repartiţiei statistice în jurul mediei aritmetice.

Amplitudinea statistică reprezintă diferenţa dintre cea mai mare valoare ($x_{max.}$) şi cea mai mică valoare ($x_{min.}$) a repartiţiei statistice. Prin urmare,

$$\bar{R} = x_{max} - x_{min}$$

Abaterea medie absolută statistică se defineşte prin

$$\bar{L} = \frac{1}{n} \sum_{i=1}^{k} |x_i - \bar{x}| n_i$$

Dispersia statistică se defineşte prin

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

pentru o repartiţie statistică de tipul I şi

$$s^2 = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i$$

pentru o repartiţie statistică de tipul II .

Evident, dispersia statistică s^2 este tocmai momentul centrat statistic de ordinul doi $\bar{\mu}_2$. De multe ori, este mai avantajos, a folosi drept dispersie statistică, mărimea

$$s_*^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

pentru o repartiţie statistică de tipul I şi

$$s_*^2 = \frac{1}{n-1} \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i$$

pentru o repartiţie statistică de tipul II.

Abaterea medie pătratică statistică se defineşte prin

$$s = \sqrt{s^2}$$

respectiv

$$s_* = \sqrt{s_*^2}$$

Coeficientul de variaţie statistic se defineşte prin

$$\bar{v} = \frac{s}{\bar{x}}$$

Asimetria poligonului frecvenţelor se apreciază cu ajutorul coeficientului de asimetrie statistic, definit prin

$$\bar{\gamma}_1 = \frac{\bar{\mu}_3}{s^3}$$

Dacă $\bar{\gamma}_1 > 0$, majoritatea valorilor repartiţiei statistice se află la dreapta mediei aritmetice $\bar{x}$. Din acest motiv, se spune că repartiţia statistică are asimetrie pozitivă.

Dacă $\bar{\gamma}_1 < 0$, majoritatea valorilor repartiţiei statistice se află la stînga mediei aritmetice $\bar{x}$. Din acest motiv, se spune că repartiţia statistică are asimetrie negativă. Dacă $\bar{\gamma}_1 = 0$ se spune că repartiţia statistică este simetrică.

Prin repartiţie statistică normală se înţelege repartiţia statistică asociată unei selecţii efectuate dintr-o colectivitate a cărei repartiţie probabilistică este o repartiţie normală. Coeficientul de asimetrie $\bar{\gamma}_1$, pentru o repartiţie statistică normală, este egal cu zero.

Boltirea poligonului frecvenţelor se apreciază cu ajutorul coeficientului de exces, definit prin

$$\bar{\gamma}_2 = \frac{\bar{\mu}_4}{s^4} - 3$$

Faptul că $\bar{\gamma}_2 > 0$, indică o concentrare a frecvenţelor în jurul tendinţei centrale, comparativ cu repartiţia statistică normală, în presupunerea că cele două repartiţii statistice au aceeaşi

abatere medie pătratică statistică s . Repartiţia statistică pentru care $\bar{\gamma}_2 > 0$ se numeşte repartiţie leptocurtică . Poligonul frecvenţelor, pentru o repartiţie statistică leptocurtică, este mai înalt decît poligonul frecvenţelor pentru repartiţia statistică normală.

Faptul că $\bar{\gamma}_2 < 0$, indică o rarefiere a frecvenţelor în jurul tendinţei centrale, comparativ cu repartiţia statistică normală. Repartiţia statistică pentru care $\bar{\gamma}_2 < 0$ se numeşte repartiţie platocurtică. Poligonul frecvenţelor, pentru o repartiţie platocurtică este mai puţin înalt decît poligonul frecvenţelor pentru repartiţia statistică normală.

Dacă $\bar{\gamma}_2 = 0$, atunci repartiţia statistică se apropie de o repartiţie statistică normală. Repartiţia statistică pentru care $\bar{\gamma}_2 = 0$ se numeşte repartiţie mezocurtică.

P R O B L E M E

1. Fie X abaterea diametrului inteior al unor inele de la dimensiunea standard. In urma unei selecţii repetate, efectuată din producţia de inele, s-a obţinut repartiţia statistică

$$X : \begin{bmatrix} -4 & -3 & -2 & -1 & 0 & 2 & 3 \\ 1 & 4 & 3 & 9 & 10 & 2 & 1 \end{bmatrix}$$

Se cere:

a) poligonul frecvenţelor absolute şi relative

b) momentele iniţiale şi centrate statistice de primele patru ordine, calculate prin metoda zeroului fals

c) funcţia de frecvenţă cumulată şi valorile ei în punctele $x_1 = -3$, $x_2 = 2$

d) indicatorii de poziţie

e) indicatorii variaţiei, asimetria, excesul

2. Un grup de persoane se repartizează după înălţime, aşa cum rezultă din tabelul 1

T a b e l u l 1.

Înălţime (cm)	(150-155]	(155-160]	(160-165]	(165-170]	(170-175]	(175-180]
Număr de persoane	1	3	14	25	5	2

Se cere:

a) poligonul şi histograma frecvenţelor absolute şi relative

b) momentele iniţiale şi centrate statistice de primele patru ordine, calculate prin metoda zeroului fals şi apoi corectate cu ajutorul formulelor lui Sheppard

c) funcţia de frecvenţă cumulată şi valorile ei în punctele $x_1 = 152$, $x_2 = 168$

d) indicatorii de poziţie

e) indicatorii variaţiei, asimetria, excesul

3. Fie repartiţia statistică

$$X : \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 0{,}135 & 0{,}269 & 0{,}270 & 0{,}180 & 0{,}100 & 0{,}030 & 0{,}010 & 0{,}003 & 0{,}002 & 0{,}001 \end{bmatrix}$$

Se cere:

a) poligonul frecvenţelor relative

b) funcţia de frecvenţă cumulată

c) să se ajusteze repartiţia statistică printr-o repartiţie probabilistică Poisson de parametru λ .

d) să se ajusteze repartiţia statistică printr-o repartiţie probabilistică binomială de parametri $n = 10$, $p = \frac{\lambda}{10}$

e) să se compare repartiţiile ajustate cu repartiţie statistică folosind indicatorul statistic

$$\bar{D} = \max_{0 \leq i \leq 9} \{|\bar{F}(i) - F(i)|\}$$

unde $\bar{F}(x)$ este funcţia de frecvenţă cumulată a repartiţiei statistice şi $F(x)$ este funcţia de repartiţie a repartiţiei ajustate.

4. Fie repartiţia statistică dată de tabelul 2

T a b e l u l 2.

Interval de grupare	(-4,-3]	(-3,-2]	(-2,-1]	(-1,0]	(0,1]	(1,2]	(2,3]	(3,4]
Frecvenţa relativă	0,011	0,051	0,140	0,270	0,244	0,172	0,090	0,022

Se cere:

a) poligonul şi histograma frecvenţelor relative

b) funcţia de frecvenţă cumulată

c) să se ajusteze repartiţia statistică printr-o repartiţie probabilistică normală de parametri m şi σ^2.

Scanned with OKEN Scanner

Capitolul III

TEORIA SELECȚIEI

Problema teoriei selecţiei constă în determinarea caracteristicilor funcţionale (funcţia de repartiţie, funcţia caracteristică, etc.) şi numerice (momente, cvantile, etc.) ale unei colectivităţi statistice pe baza selecţiilor efectuate din colectivitatea considerată. Pentru a putea rezolva această problemă cu ajutorul teoriei probabilităţilor este necesar a fundamenta probabilistic noţiunile de colectivitate statistică, selecţie, caracteristică a unei colectivităţi statistice.

§ 1. FUNDAMENTELE TEORIEI SELECŢIEI.

Un individ oarecare al unei colectivităţi statistice îl vom nota cu ω . Fie Ω mulţimea tuturor indivizilor unei colectivităţi şi fie $\mathcal{K}$ un corp borelian de părţi ale lui Ω .

Prin colectivitate statistică se înţelege cuplul $\{\Omega, \mathcal{K}\}$. Prin urmare, modelul teoretic al colectivităţii statistice îl constituie cîmpul de evenimente. Pe cîmpul borelian de evenimente $\mathcal{K}$ se presupune definită o probabilitate P cunoscută sau necunoscută. Dacă volumul colectivităţii este finit, atunci $\mathcal{K}$ coincide cu mulţimea tuturor părţilor lui Ω , iar P este o repartiţie probabilistică discretă uniformă pe $\mathcal{K}$.

Fie

$$\Omega^{(n)} = \underbrace{\Omega x \ldots x \Omega}_{\text{de n ori}} \quad , \quad \mathcal{K}^{(n)} = \underbrace{\mathcal{K} x \ldots x \mathcal{K}}_{\text{de n ori}}$$

şi probabilitatea

$$P^{(n)} = \underbrace{P \; x \ldots x \; P}_{\text{de n ori}}$$

pe $\mathcal{K}^{(n)}$ indusă de P . Un element al lui $\Omega^{(n)}$ este

$$\omega^{(n)} = (\omega_1, \omega_2, \ldots, \omega_n)$$

Definiţia 1. Se numeşte selecţie repetată de volum n orice element $\omega^{(n)} \in \Omega^{(n)}$.

Cuplul $\{\Omega^{(n)}, \mathcal{K}^{(n)}\}$ se numeşte spaţiul selecţiilor repetate de volum n .

Definiţia 2. Se numeşte caracteristică a colectivităţii $\{\Omega, \mathcal{K}\}$ orice aplicaţie $X : \Omega \longrightarrow R$, cu proprietatea că, $X^{-1}(B) \in \mathcal{K}$, oricare ar fi mulţimea boreliană B a lui R .

Deci, dacă colectivitatea statistică este considerată ca un cîmp de evenimente, orice caracteristică a colectivităţii este o variabilă aleatoare. De obicei, $\mathcal{K}$ este cel mai mic corp borelian în raport cu care X este măsurabilă. Caracteristicile funcţionale şi numerice ale lui X se numesc caracteristici ale colectivităţii sau caracteristici teoretice.

Fie X o caracteristică urmărită la indivizii colectivităţii $\{\Omega, \mathcal{K}\}$ şi fie $\{\Omega^{(n)}, \mathcal{K}^{(n)}\}$ spaţiul selecţiilor repetate de volum n . Considerăm aplicaţiile

$$X_i : \Omega^{(n)} \longrightarrow R$$

definite prin

$$X_i(\omega^{(n)}) = X(\omega_i) , \quad 1 \leqslant i \leqslant n .$$

Teorema 1. Aplicaţiile X_i, $1 \leqslant i \leqslant n$, sînt variabile aleatoare definite pe $\{\Omega^{(n)}, \mathcal{K}^{(n)}\}$.

Demonstraţie. Fie B o mulţime boreliană a lui R. Pentru $1 \leqslant i \leqslant n$, avem

$$X_i^{-1}(B) = \{\omega^{(n)} : X_i(\omega^{(n)}) \in B\} = \{\omega^{(n)} : X(\omega_i) \in B\} =$$

$$= \Omega \times \ldots \times \{\omega_i : X(\omega_i) \in B\} \times \ldots \times \Omega =$$

$$= \Omega \times \ldots \times X^{-1}(B) \times \ldots \times \Omega \in \mathcal{K}^{(n)} .$$

Definiţia 3. Se numesc variabile aleatoare de selecţie repetată de volum n , ale caracteristicii X, variabilele aleatoare $X_1, X_2, \ldots, X_n$, definite pe $\{\Omega^{(n)}, \mathcal{K}^{(n)}\}$, prin

$$X_i(\omega^{(n)}) = X(\omega_i) , \quad 1 \leqslant i \leqslant n .$$

Teorema 2. Variabilele aleatoare de selecţie X_i, $1 \leqslant i \leqslant n$,

au aceeaşi funcţie de repartiţie ca şi caracteristica X şi sînt independente.

Demonstraţie. Fie $B = (-\infty, x) \subset R$. Din teorema 1 rezultă că pentru $1 \leq i \leq n$, avem

$$X_i^{-1}(-\infty, x) = \Omega \times \ldots \times X^{-1}(-\infty, x) \times \ldots \times \Omega \in \mathcal{K}^{(n)}$$

Ţinînd seama de definiţia funcţiei de repartiţie şi a probabilităţii $P^{(n)}$ se obţine

$$P^{(n)}(X_i^{-1}(-\infty, x)) = P(\Omega) \ldots P(X^{-1}(-\infty, x)) \ldots P(\Omega) = P(X^{-1}(-\infty), x)) = F(x)$$

unde $F(x)$ este funcţia de repartiţie a lui X . Deci, toate variabilele aleatoare de selecţie X_i, $1 \leq i \leq n$, au funcţia de repartiţie $F(x)$. Pentru a demonstra independenţa, vom arăta că dacă B_i , $1 \leq i \leq n$, sînt mulţimi boreliene ale lui R , atunci

$$P^{(n)}\left(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\right) = \prod_{i=1}^{n} P^{(n)}(X_i^{-1}(B_i))$$

Intr-adevăr, avem

$$P^{(n)}\left(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\right) = P^{(n)}((\omega_1, \ldots, \omega_n) : X(\omega_1) \in B_1, \ldots, X(\omega_n) \in B_n) =$$

$$= \prod_{i=1}^{n} P(\omega_i : X(\omega_i) \in B_i) = \prod_{i=1}^{n} P^{(n)}(X_i^{-1}(B_i)) .$$

Din teorema 2 rezultă că toate caracteristicile funcţionale şi numerice ale variabilelor aleatoare de selecţie X_i, $1 \leq i \leq n$, sînt egale cu caracteristicile corespunzătoare ale lui X . Din acest motiv, variabilele aleatoare de selecţie repetată de volum n se numesc exemplare ale lui X .

Definiţia 4. Se numeşte vector de selecţie repetată de volum n , al caracteristicii X , vectorul aleator $Y : \Omega^{(n)} \longrightarrow R^n$, definit prin $Y(\omega^{(n)}) = (X_1(\omega^{(n)}), \ldots, X_n(\omega^{(n)}))$

Definiţia 5. Se numesc valori de selecţie repetată de volum n, ale caracteristicii X, componentele vectorului aleator $Y(\omega^{(n)})$ pentru $\omega^{(n)}$ fixat.

Dacă notăm $L^n = Y(\Omega^{(n)}) \subseteq R^n$ mulţimea valorilor vectorului aleator Y , atunci orice punct $(x_1,x_2,\dots,x_n) \in L^n$ este o valoare de selecţie. Mulţimea L^n a valorilor de selecţie constituie spaţiul valorilor de selecţie. In cele ce urmează, printr-o selecţie repetată de volum n, vom înţelege valoarea de selecţie $(x_1,x_2,\dots,x_n)$, unde $x_i = X_i(\omega^{(n)})$, $1 \leqslant i \leqslant n$. Modelul teoretic al selecţiei repetate îl constituie urna cu bile revenite.

In cazul selecţiei nerepetate variabilele aleatoare de selecţie $X_1,X_2,\dots,X_n$ au aceeaşi funcţie de repartiţie ca şi X, dar nu mai sînt independente. Modelul teoretic al selecţiei nerepetate îl constituie urna cu bile nerevenite. In cele ce urmează vom considera numai selecţii repetate.

Definiţia 6. Se numeşte statistică (funcţie de selecţie) o funcţie reală, măsurabilă Borel, de variabilele aleatoare de selecţie $X_1,X_2,\dots,X_n$.

Cum orice funcţie reală, măsurabilă Borel, de variabile aleatoare este o variabilă aleatoare, rezultă că o statistică este o variabilă aleatoare. O statistică o vom nota prin $S(X_1,\dots,X_n)$ sau $S(X,n)$. In cele ce urmează vom studia următoarele statistici: funcţia de repartiţie de selecţie, momentele iniţiale de selecţie şi momentele centrate de selecţie.

§ 2. FUNCŢIA DE REPARTIŢIE DE SELECŢIE

Presupunem că valorile x_i , $1 \leqslant i \leqslant n$, unde $x_i = X_i(\omega^{(n)})$ sînt ordonate în ordine nedescrescătoare.

Definiţia 1. Se numeşte funcţie de repartiţie de selecţie aplicaţia $F_n^* : R \times \Omega^{(n)} \longrightarrow R$, care pentru fiecare $\omega^{(n)}$ fixat ia valorile

$$F_n^*(x, \omega^{(n)}) = \begin{cases} 0 & x \leqslant x_1 \\ \frac{k}{n} & x_k < x \leqslant x_{k+1}, \ (k = 1,2,\ldots,n-1) \\ 1 & x > x_n \end{cases}$$

Se observă că, pentru $\omega^{(n)}$ fixat, funcţia de repartiţie de selecţie coincide cu funcţia de frecvenţă cumulată definită în §5, cap.II.

Teorema 1. Pentru x fixat, funcţia de repartiţie de selecţie $F_n^*(x, \omega^{(n)})$ este o variabilă aleatoare pe $\Omega^{(n)}$, avînd tabloul de repartiţie

$$\begin{bmatrix} \frac{k}{n} \\ C_n^k [F(x)]^k [1 - F(x)]^{n-k} \end{bmatrix}_{0 \leqslant k \leqslant n}$$

Demonstraţie. A se vedea Reischer Corina, Sâmboan G., R. Theodorescu [15] .

Se consideră pentru x fixat, evenimentul

$$A_x = \{ \omega^{(n)} : X_i(\omega^{(n)}) < x \}, \ 1 \leqslant i \leqslant n .$$

Potrivit definiţiei 1 şi definiţiei funcţiei de repartiţie teoretice, se observă că , $F_n^*(x, \omega^{(n)})$ este frecvenţa relativă a lui A_x în n probe independente, iar $F(x)$ este probabilitatea lui A_x .

Folosind teorema 1 din acest paragraf şi teorema 1, §2, cap. I, se obţine că, pentru x fixat, au loc relaţiile

$$M[F_n^*(x, \omega^{(n)})] = F(x)$$

$$D^2[F_n^*(x, \omega^{(n)})] = \frac{F(x)[1 - F(x)]}{n}$$

Teorema 2. Pentru fiecare $x \in R$ şirul funcţiilor de repartiţie de selecţie $(F_n^*(x, \omega^{(n)}))_{n \in N^*}$ converge în probabilitate către $F(x)$.

Demonstraţie. Teorema rezultă din teorema lui Bernoulli, § 1, cap.I şi definiţiile funcţiilor $F_n^*(x, \omega^{(n)})$ şi $F(x)$.

Teorema 3. Pentru fiecare $x \in R$ şirul funcţiilor de repartiţie $(F_n^*(x, \omega^{(n)})_{n \in N^*}$ converge aproape sigur către $F(x)$.

Scanned with OKEN Scanner

Demonstraţie. Teorema rezultă din teorema lui Borel, § 1, cap.I şi definiţiile funcţiilor $F_n^*(x, \omega^{(n)})$ şi $F(x)$.

Asupra relaţiei dintre funcţia de repartiţie de selecţie $F_n^*(x, \omega^{(n)})$ şi funcţia de repartiţie teoretică $F(x)$, vom enunţa două teoreme.

Teorema 4. (V.I. Glivenko)

Dacă $F(x)$ este funcţia de repartiţie teoretică, $F_n^*(x, \omega^{(n)})$ este funcţia de repartiţie de selecţie şi

$$D_n(\omega^{(n)}) = \sup_{x \in R} |F_n^*(x, \omega^{(n)}) - F(x)|$$

atunci $$P^{(n)}(\omega^{(n)} : \lim_{n\to\infty} D_n(\omega^{(n)}) = 0) = 1$$

Demonstraţie. A se vedea Mihoc Gh., Ciucu G., Craiu V. [10].

Teorema lui Glivenko, afirmă că, şirul de statistici $(D_n(\omega^{(n)}))_{n\in N^*}$ converge aproape sigur către zero. Această teoremă constituie justificarea teoretică a folosirii metodei selecţiei ca metodă de cercetare a unei colectivităţi statistice.

Teorema 5. (A.N. Kolmogorov)

Dacă $F(x)$ este funcţia de repartiţie teoretică, continuă, $F_n^*(x, \omega^{(n)})$ este funcţia de repartiţie de selecţie şi

$$D_n(\omega^{(n)}) = \sup_{x \in R} |F_n^*(x, \omega^{(n)}) - F(x)|$$

atunci

$$\lim_{n\to\infty} P^{(n)}(\omega^{(n)} : D_n(\omega^{(n)}) < \frac{\lambda}{\sqrt{n}}) = \begin{cases} K(\lambda) & \lambda > 0 \\ 0 & \lambda \leq 0 \end{cases}$$

unde $$K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$$

Demonstraţie. A se vedea Mihoc Gh., Ciucu G., Craiu V [10].

Valorile funcţiei $K(\lambda)$, numită funcţia lui Kolmogorov, se află în tabelul X din anexă.

§ 3. MEDIA DE SELECŢIE

Definiţie. Se numeşte **medie de selecţie** statistica

$$m(X,n) = \int_{-\infty}^{+\infty} x \, dF_n^*(x, \omega^{(n)})$$

Ţinînd seama de proprietăţile lui $F_n^*(x, \omega^{(n)})$, se obţine

$$m(X,n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

unde X_i , $1 \leqslant i \leqslant n$, sînt variabilele aleatoare de selecţie repetată de volum n , ale caracteristicii X .

Teorema 1. Dacă colectivitatea admite medie, atunci media mediei de selecţiei este egală cu media colectivităţii.

Demonstraţie. Din faptul că X_i , $1 \leqslant i \leqslant n$, au aceeaşi funcţie de repartiţie ca şi X , rezultă că , $M(X_i) = M(X)$, $1 \leqslant i \leqslant n$. Prin urmare, avem

$$M[m(X,n)] = M\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} M(X_i) = \frac{1}{n} \sum_{i=1}^{n} M(X) = M(X) .$$

Teorema 2. Dacă colectivitatea admite dispersie, atunci dispersia mediei de selecţiei este de n ori mai mică decît dispersia colectivităţii, unde n este volumul selecţiei.

Demonstraţie. Din faptul că X_i, $1 \leqslant i \leqslant n$, sînt independente şi au aceeaşi funcţie de repartiţie ca şi X , rezultă că

$$D^2[m(X,n)] = D^2\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} D^2(X) = \frac{D^2(X)}{n}$$

Teorema 3. Media de selecţie este translativă.

Demonstraţie. Fie c un număr real. Atunci

$$m(X + c,n) = \frac{1}{n} \sum_{i=1}^{n} (X_i + c) = \frac{1}{n} \sum_{i=1}^{n} X_i + c = m(X,n) + c .$$

Observaţie. Valoarea mediei de selecţie $m(X,n)$ pentru selecţia $(x_1, \ldots, x_n)$ fixată este $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Teorema 4. Media de selecţie converge în probabilitate către media colectivităţii.

Demonstraţie. Se disting două cazuri :

a) Dacă colectivitatea admite medie finită, atunci teorema 4 rezultă din teorema 1 a acestui paragraf şi teorema lui Hincin, §1, cap.I.

b) Dacă colectivitatea admite momente iniţiale de primele două ordine, finite, atunci ea admite şi dispersie finită. Aplicînd inegalitatea lui Cebîşev statisticii $m(X,n)$, se obţine că, pentru orice $\varepsilon > 0$, are loc

$$0 \leqslant P(|m(X,n) - M(m(X,n))| \geqslant \varepsilon) \leqslant \frac{D^2[m(X,n)]}{\varepsilon^2} .$$

Ţinînd seama de teoremele 1 şi 2 ale acestui paragraf, ultima relaţie devine

$$0 \leqslant P(|m(X,n) - M(X)| \geqslant \varepsilon) \leqslant \frac{D^2(X)}{n\varepsilon^2}$$

Făcînd $n \longrightarrow \infty$ în ultima relaţie, se obţine că, oricare ar fi $\varepsilon > 0$, are loc

$$\lim_{n\to\infty} P(|m(X,n) - M(X)| \geqslant \varepsilon) = 0$$

ceea ce înseamnă, conform definiţiei 14, §1, cap.I, că media de selecţie converge în probabilitate către media colectivităţii.

Teorema 5. Dacă colectivitatea admite

$$M(X) = m \ , \ D^2(X) = \sigma^2 \ , \ M(|X - m|^3) = H^3$$

finite, atunci statistica

$$S(X,n) = \frac{m(X,n) - m}{\frac{\sigma}{\sqrt{n}}}$$

urmează asimptotic repartiţia $N(0,1)$.

Demonstraţie. Vom verifica îndeplinirea condiţiilor teoremei centrală limită a lui Leapunov din §1, cap.I. Fie $Y_i = \frac{X_i}{n}$, $1 \leqslant i \leqslant n$. Atunci, avem

$$m_i = M(Y_i) = M(\frac{X_i}{n}) = \frac{1}{n} M(X_i) = \frac{m}{n}$$

$$D_i^2 = D^2(Y_i) = D^2(\frac{X_i}{n}) = \frac{1}{n^2} D^2(X_i) = \frac{\sigma^2}{n^2}$$

$$H_i^3 = M(|Y_i - m_i|^3) = M(|\frac{X_i}{n} - \frac{m}{n}|^3) = \frac{1}{n^3} M(|X_i - m|^3) = \frac{H^3}{n^3}$$

pentru $1 \leqslant i \leqslant n$. De aici, rezultă

$$S_n = \sqrt{D_1^2 + \ldots + D_n^2} = \frac{\sigma}{\sqrt{n}}, \quad K_n = \sqrt[3]{H_1^3 + \ldots + H_n^3} = \frac{H}{n^{2/3}} .$$

Prin urmare,

$$\lim_{n\to\infty} \frac{K_n}{S_n} = \lim_{n\to\infty} \frac{H}{\sigma n^{1/6}} = 0 .$$

Conform teoremei lui Leapunov, statistica

$$S(X,n) = \frac{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} M(Y_i)}{S_n} = \frac{m(X,n) - m}{\frac{\sigma}{\sqrt{n}}}$$

urmează asimptotic repartiţia $N(0,1)$.

Teorema 6. Dacă repartiţia colectivităţii este o repartiţie Poisson de parametru λ , atunci media de selecţie urmează o repartiţie a cărei funcţie caracteristică este

$$\Phi(t) = e^{n\lambda(e^{i\,t/n} - 1)}$$

Demonstraţie. Conform teoremei 6, §2, cap.I, funcţia caracteristică a colectivităţii este

$$\varphi_X(t) = e^{\lambda(e^{it} - 1)}$$

Ţinînd seama că X_i, $1 \leqslant i \leqslant n$, sînt independente şi au aceeaşi funcţie de repartiţie ca şi X, rezultă

$$\Phi(t) = M\left[e^{itm(X,n)}\right] = M\left[e^{i\frac{t}{n}\sum_{k=1}^{n} X_k}\right] = \prod_{k=1}^{n} M\, e^{i\frac{t}{n}X_k} =$$

$$= \prod_{k=1}^{n} \varphi_{X_k}(\frac{t}{n}) = \prod_{k=1}^{n} \varphi_X(\frac{t}{n}) = \prod_{k=1}^{n} e^{\lambda(e^{i\frac{t}{n}} - 1)} = e^{n\lambda(e^{i\frac{t}{n}} - 1)} .$$

§ 4. MOMENTE INIŢIALE DE SELECŢIE.

Definiţie. Se numeşte moment iniţial de selecţie de ordin k, (k = 0,1,2,...,), statistica

$$a_k(X,n) = \int_{-\infty}^{+\infty} x^k dF_n^*(x, \omega^{(n)})$$

Ţinînd seama de proprietăţile lui $F_n^*(x, \omega^{(n)})$, se obţine

$$a_k(X,n) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

unde X_i , $1 \leqslant i \leqslant n$, sînt variabilele aleatoare de selecţie repetată de volum n, ale caracteristicii X . Evident, momentul iniţial de selecţie de ordinul întîi este media de selecţie.

Teorema 1. Dacă colectivitatea admite moment iniţial de ordin k, atunci media momentului iniţial de selecţie de ordin k este egală cu momentul iniţial de ordin k al colectivităţii.

Demonstraţie. Din faptul că X_i, $1 \leqslant i \leqslant n$, au aceeaşi funcţie de repartiţie ca şi X , rezultă că , $M(X_i^k) = M(X^k) = \alpha_k(X)$, $1 \leqslant i \leqslant n$. Prin urmare avem

$$M[a_k(X,n)] = M\left[\frac{1}{n} \sum_{i=1}^{n} X_i^k\right] = \frac{1}{n} \sum_{i=1}^{n} M(X_i^k) =$$

$$= \frac{1}{n} \sum_{i=1}^{n} M(X^k) = \frac{1}{n} \sum_{i=1}^{n} \alpha_k(X) = \alpha_k(X) .$$

Teorema 2. Dacă colectivitatea admite momente iniţiale de ordinele k şi 2k , atunci dispersia momentului iniţial de selecţie de ordin k este de n ori mai mică decît dispersia lui X^k , unde n este volumul selecţiei.

Demonstraţie. Intrucît

$$D^2(X^k) = M(X^{2k}) - M^2(X^k) = \alpha_{2k}(X) - \alpha_k^2(X)$$

avem de demonstrat că dispersia lui $a_k(X,n)$ este egală cu

$$\frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n} .$$

Din independenţa variabilelor aleatoare de selecţie X_i, rezultă independenţa variabilelor aleatoare X_i^k, $1 \leqslant i \leqslant n$. Ţinînd seama de acest lucru şi de faptul că X_i, $1 \leqslant i \leqslant n$, au aceeaşi repartiţie ca şi X, rezultă

$$D^2[a_k(X,n)] = D^2\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i^k) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}[M(X_i^{2k}) - M^2(X_i^k)] = \frac{1}{n^2}\sum_{i=1}^{n}[M(X^{2k}) - M^2(X^k)] =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}[\alpha_{2k}(X) - \alpha_k^2(X)] = \frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n} .$$

Observaţie. Valoarea momentului iniţial de selecţie $a_k(X,n)$ pentru selecţia $(x_1, x_2, \ldots, x_n)$ fixată este

$$\bar{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k .$$

Teorema 3. Momentul iniţial de selecţie de ordin k, converge în probabilitate către momentul iniţial de ordin k al colectivităţii.

Demonstraţie. Dacă se presupune că, colectivitatea admite momente iniţiale de ordinele k şi 2k, finite, atunci aplicînd statisticii $a_k(X,n)$ inegalitatea lui Cebîşev se obţine că, pentru orice $\varepsilon > 0$, are loc

$$0 \leqslant P(|a_k(X,n) - M[a_k(X,n)]| \geqslant \varepsilon) \leqslant \frac{D^2[a_k(X,n)]}{\varepsilon^2} .$$

Ţinînd seama de teoremele 1 şi 2 din acest paragraf, ultima relaţie devine

$$0 \leqslant P(|a_k(X,n) - \alpha_k(X)| \geqslant \varepsilon) \leqslant \frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n\varepsilon^2}$$

de unde, pentru $n \longrightarrow \infty$, se obţine că, oricare ar fi $\varepsilon > 0$, are loc

$$\lim_{n\to\infty} P(|a_k(X,n) - \alpha_k(X)| \geq \varepsilon) = 0$$

Teorema 4. Dacă colectivitatea admite

$$\alpha_k(X) = M(X^k)$$
$$\alpha_{2k}(X) = M(X^{2k})$$
$$H^3 = M(|X^k - \alpha_k(X)|^3)$$

finite, atunci statistica

$$S(X,n) = \frac{a_k(X,n) - \alpha_k(X)}{\sqrt{\frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n}}}$$

urmează asimptotic repartiţia $N(0,1)$.

Demonstraţie. Vom verifica îndeplinirea condiţiilor teoremei centrală limită a lui Leapunov din §1, cap.I. Fie

$Y_i = \frac{X_i^k}{n}$, $1 \leq i \leq n$, Atunci, avem

$$m_i = M(Y_i) = M(\frac{X_i^k}{n}) = \frac{1}{n} M(X_i^k) = \frac{1}{n} \alpha_k(X_i) = \frac{\alpha_k(X)}{n}$$

$$D_i^2 = D^2(Y_i) = D^2(\frac{X_i^k}{n}) = \frac{1}{n^2} D^2(X_i^k) = \frac{1}{n^2}\left[M(X_i^{2k}) - M^2(X_i^k)\right] =$$

$$= \frac{1}{n^2}\left[\alpha_{2k}(X_i) - \alpha_k(X_i)\right] = \frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n^2}$$

$$H_i^3 = M(|Y_i - m_i|^3) = M(|\frac{X_i^k}{n} - \frac{\alpha_k(X)}{n}|^3) =$$

$$= \frac{1}{n^3} M(|X_i^k - \alpha_k(X)|^3) = \frac{1}{n^3} M(|X^k - \alpha_k(X)|^3) = \frac{H^3}{n^3}$$

pentru $1 \leq i \leq n$. De aici, rezultă

$$S_n = \sqrt{D_1^2 + \ldots + D_n^2} = \sqrt{\frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n}}$$
$$K_n = \sqrt[3]{H_1^3 + \ldots + H_n^3} = \frac{H}{n^{2/3}}$$

Prin urmare

$$\lim_{n\to\infty} \frac{K_n}{S_n} = \lim_{n\to\infty} \frac{H}{n^{1/6}\sqrt{\alpha_{2k}(X) - \alpha_k^2(X)}} = 0 .$$

Conform teoremei lui Leapunov, statistica

$$S(X,n) = \frac{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} M(Y_i)}{S_n} = \frac{a_k(X,n) - \alpha_k(X)}{\sqrt{\frac{\alpha_{2k}(X) - \alpha_k^2(X)}{n}}}$$

urmează asimptotic repartiţia $N(0,1)$.

§ 5. DISPERSIA DE SELECŢIE .

Definiţie. Se numeşte dispersie de selecţie statistica

$$d^2(X,n) = \int_{-\infty}^{+\infty} [x - m(X,n)]^2 dF_n^*(x, \omega^{(n)})$$

Ţinînd seama de proprietăţile lui $F_n^*(x, \omega^{(n)})$, se obţine

$$d^2(X,n) = \frac{1}{n} \sum_{i=1}^{n} [X_i - m(X,n)]^2$$

unde X_i, $1 \leq i \leq n$, sînt variabilele aleatoare de selecţie repetată de volum n , ale caracteristicii X .

Teorema 1. Dispersia de selecţie se exprimă cu ajutorul momentelor iniţiale de selecţie, prin formula

$$d^2(X,n) = a_2(X,n) - a_1^2(X,n)$$

Demonstraţie. Ţinînd seama de definiţia lui $d^2(X,n)$, avem

$$d^2(X,n) = \frac{1}{n} \sum_{i=1}^{n} [X_i - m(X,n)]^2 = \frac{1}{n} \sum_{i=1}^{n} [X_i^2 - 2m(X,n)X_i + m^2(X,n)] =$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - 2m(X,n) \cdot \frac{1}{n} \sum_{i=1}^{n} X_i + m^2(X,n) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - 2m^2(X,n) +$$

$$+ m^2(X,n) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - m^2(X,n) = a_2(X,n) - a_1^2(X,n) \quad .$$

Observaţia 1. Dispersia de selecţie se exprimă în funcţie de momentele iniţiale de selecţie prin aceeaşi formulă prin care se exprimă dispersia teoretică în funcţie de momentele iniţiale teoretice.

Teorema 2. Dispersia de selecţie este invariantă la translaţie.

Demonstraţie. Fie c un număr real. Folosind teorema 3, §3, cap. III, rezultă

$$d^2(X + c,n) = \frac{1}{n}\sum_{i=1}^{n}[X_i + c - m(X + c,n)]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[X_i + c - m(X,n) - c]^2 = \frac{1}{n}\sum_{i=1}^{n}[X_i - m(X,n)]^2 = d^2(X,n) .$$

Observaţia 2. Potrivit teoremei 2, dacă se consideră caracteristica

$$Y = X - M(X)$$

atunci $d^2(Y,n) = d^2(X,n)$. Evident, $M(Y) = 0$. Din aceste motive, în teoremele ce se referă la dispersia de selecţie se poate face de la început presupunerea că $M(X) = 0$, fără a restrînge generalitatea acestora.

Teorema 3. Dacă colectivitatea admite dispersie, atunci media dispersiei de selecţiei este

$$M[d^2(X,n)] = \frac{n-1}{n} D^2(X)$$

unde n este volumul selecţiei.

Demonstraţie. Presupunem că $M(X) = 0$. Din această presupunere rezultă

$$\mu_k(X) = M[X - M(X)]^k = M(X^k) = \alpha_k(X) , \quad k = 0,1,2,...$$

şi

$$\mu_k[m(X,n)] = M\{m(X,n) - M[m(X,n)]\}^k = M[m(X,n) - M(X)]^k =$$

$$= M[m^k(X,n)] = \alpha_k[m(X,n)] , \quad k = 0,1,2,...$$

Folosind teorema 1 din acest paragraf, teorema 1, §4, cap. III şi teorema 2, § 3, cap. III, se obţine

$$M[d^2(X,n)] = M[a_2(X,n) - a_1^2(X,n)] = M[a_2(X,n)] - M[a_1^2(X,n)] =$$

$$= \alpha_2(X) - \alpha_2[a_1(X,n)] = \alpha_2(X) - \alpha_2[m(X,n)] = \mu_2(X) -$$

$$- \mu_2[m(X,n)] = D^2(X) - \frac{D^2(X)}{n} = \frac{n-1}{n} D^2(X) .$$

Teorema 4. Dacă colectivitatea admite momente centrate de ordinele doi şi patru, atunci dispersia dispersiei de selecţie este

$$D^2[d^2(X,n)] = \frac{\mu_4(X)-\mu_2^2(X)}{n} - \frac{2[\mu_4(X)-2\mu_2^2(X)]}{n^2} + \frac{\mu_4(X)-3\mu_2^2(X)}{n^3}.$$

Demonstrație. Presupunem că $M(X) = 0$. Din această presupunere, așa după cum s-a arătat în demonstrația teoremei 3 rezultă

$$\mu_k(X) = \alpha_k(X) \text{ și } \mu_k[m(X,n)] = \alpha_k[m(X,n)], \quad k = 0,1,2,\ldots$$

Ținînd seama de definiția 6, §1, cap.I și teorema 3 din acest paragraf, rezultă

$$D^2[d^2(X,n)] = M\{[d^2(X,n)]^2\} - M^2[d^2(X,n)] =$$

$$= M\{[d^2(X,n)]^2\} - \frac{(n-1)^2}{n^2}\mu_2^2(X).$$

Conform teoremei 1 din acest paragraf, avem

$$M\{[d^2(X,n)]^2\} = M[a_2(X,n) - m^2(X,n)]^2 = M[a_2^2(X,n)] -$$

$$- 2M[a_2(X,n)m^2(X,n)] + M[m^4(X,n)].$$

Calculînd fiecare termen din membrul drept al ultimei relații obținem

$$M[a_2^2(X,n)] = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right]^2 = \frac{1}{n^2} M\left[\sum_{i=1}^{n} X_i^4 + 2\sum_{j\neq k} X_j^2 X_k^2\right] =$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} M(X_i^4) + 2\sum_{j\neq k} M(X_j^2)M(X_k^2)\right] = \frac{1}{n^2}\left[\sum_{i=1}^{n} M(X^4) + 2C_n^2 M^2(X^2)\right] =$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} \alpha_4(X) + n(n-1)\alpha_2^2(X)\right] = \frac{\alpha_4(X)-(n-1)\alpha_2^2(X)}{n} =$$

$$= \frac{\mu_4(X) + (n-1)\mu_2^2(X)}{n}$$

$$M[a_2(X,n)m^2(X,n)] = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] =$$

$$= \frac{1}{n^3} M\left[\sum_{i=1}^{n} X_i^2\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{j\neq k} X_j X_k\right)\right] =$$

$$= \frac{1}{n^3} M\left[\sum_{i=1}^{n} X_i^4 + 2\sum_{j\neq k} X_j^2 X_k^2 + \sum_{i,j\neq k} X_i^2 X_j X_k\right] =$$

$$= \frac{1}{n^3}\left[\sum_{i=1}^{n} M(X_i^4) + 2\sum_{j\neq k} M(X_j^2)M(X_k^2)\right] =$$

$$= \frac{1}{n^3}\left[\sum_{i=1}^{n} M(X^4) + 2C_n^2 M^2(X^2)\right] =$$

$$= \frac{1}{n^3}\left[nM(X^4) + n(n-1)M^2(X^2)\right] =$$

$$= \frac{\alpha_4(X) + (n-1)\alpha_2^2(X)}{n^2} = \frac{\mu_4(X) + (n-1)\mu_2^2(X)}{n^2}$$

$$M\left[m^4(X,n)\right] = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^4 = \frac{1}{n^4} M\left[\sum_{\substack{k_1,\ldots,k_n \\ k_1+\ldots+k_n=4}} \frac{4!}{k_1!\ldots k_n!} X_1^{k_1}\ldots X_n^{k_n}\right] =$$

$$= \frac{1}{n^4}\left[n\,\alpha_4(X) + 3n(n-1)\alpha_2^2(X)\right] = \frac{\mu_4(X)+3(n-1)\mu_2^2(X)}{n^3}$$

Prin urmare

$$D^2\left[d^2(X,n)\right] = \frac{\mu_4(X)+(n-1)\mu_2^2(X)}{n} - \frac{2\left[\mu_4(X)+(n-1)\mu_2^2(X)\right]}{n^2} +$$

$$+ \frac{\mu_4(X) + 3(n-1)\mu_2^2(X)}{n^3} - \frac{n^2 - 2n + 3}{n^2}\mu_2^2(X) =$$

$$= \frac{\mu_4(X) - \mu_2^2(X)}{n} - \frac{2\left[\mu_4(X) - 2\mu_2^2(X)\right]}{n^2} + \frac{\mu_4(X) - 3\mu_2^2(X)}{n^3}$$

Observaţia 3. Valoarea lui $d^2(X,n)$ pentru selecţia (x_1,x_n) fixată este $\bar{\mu}_2 = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Uneori drept dispersie de selecţie se consideră statistica

$$d_*^2(X,n) = \frac{1}{n-1}\sum_{i=1}^{n}\left[X_i - m(X,n)\right]^2$$

Evident, are loc relaţia

$$d_*^2(X,n) = \frac{n}{n-1}\, d^2(X,n) .$$

Teorema 5. Dacă colectivitatea admite dispersie, atunci media dispersiei de selecţie $d_*^2(X,n)$ este egală cu dispersia colectivităţii .

Demonstrație. Ținînd seama de formula de legătură între $d^2(X,n)$ și $d^2_*(X,n)$ și de teorema 3 a acestui paragraf, avem

$$M\left[d^2_*(X,n)\right] = M\left[\frac{n}{n-1}\, d^2(X,n)\right] = \frac{n}{n-1} M\left[d^2(X,n)\right] = \frac{n}{n-1}\cdot\frac{n-1}{n} D^2(X) =$$

$$= D^2(X) .$$

Teorema 6. Dacă colectivitatea admite momente centrate de ordinele doi și patru, atunci dispersia dispersiei de selecție $d^2_*(X,n)$ se exprimă prin

$$D^2\left[d^2_*(X,n)\right] = \frac{n^2}{(n-1)^2} D^2\left[d^2(X,n)\right] =$$

$$= \frac{n^2}{(n-1)^2}\left\{\frac{\mu_4(X) - \mu_2^2(X)}{n} - \frac{2\left[\mu_4(X) - 2\mu_2^2(X)\right]}{n^2} + \frac{\mu_4(X) - 3\mu_2^2(X)}{n^3}\right\}$$

Demonstrație. Se folosește formula de legătură între $d^2(X,n)$ și $d^2_*(X,n)$ și teorema 4 din acest paragraf.

Teorema 7. Dispersia de selecție $d^2_*(X,n)$ este invariantă la translație.

Demonstrație. Analog ca teorema 2 din acest paragraf.

Observația 4. Valoarea lui $d^2_{*n}(X,n)$ pentru selecția $(x_1,\ldots,x_n)$ fixată este $s^2_* = \frac{1}{n-1}\sum_{i=1}(x_i - \bar{x})^2$.

§ 6. MOMENTE CENTRATE DE SELECȚIE.

Definiție. Se numește moment centrat de selecție de ordin k, $(k = 0,1,2,\ldots)$ statistica

$$b_k(X,n) = \int_{-\infty}^{+\infty}\left[x - m(X,n)\right]^k dF_n^*(x, \omega^{(n)})$$

Ținînd sema de proprietățile lui $F_n^*(x, \omega^{(n)})$, se obține

$$b_k(X,n) = \frac{1}{n}\sum_{i=1}^{n}\left[X_i - m(X,n)\right]^k$$

unde X_i, $1 \leq i \leq n$, sînt variabilele aleatoare de selecție repetată de volum n, ale caracteristicii X. Evident, momentul centrat de selecție de ordinul al doilea $b_2(X,n)$ este tocmai dispersia de selecție $d^2(X,n)$.

Teorema 1. Momentul centrat de selecţie de ordin k este invariant la translaţie.

Demonstraţie. Fie c un număr real. Ţinînd seama de teorema 3, §3, cap.III, avem

$$b_k(X + c,n) = \frac{1}{n}\sum_{i=1}^{n}[X_i + c - m(X+c,n)]^k = \frac{1}{n}\sum_{i=1}^{n}[X_i + c - m(X,n) - c]^k =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[X_i - m(X,n)]^k = b_k(X,n)$$

Teorema 2. Dacă colectivitatea admite moment centrat de ordin k , atunci media momentului centrat de selecţie de ordin k, este

$$M[b_k(X,n)] = \mu_k(X) + O(\frac{1}{n})$$

unde $O(\frac{1}{n})$ este o funcţie asiptotică de ordinul lui $\frac{1}{n}$.

Demonstraţie. A se vedea Mihoc G., Firescu D. [9].

Teorema 3. Dacă colectivitatea admite momente centrate de ordinele 2 , k - 1,k,k + 1,2k, atunci dispersia momentului centrat de selecţie de ordin k , este

$$D^2[b_k(X,n)] = \frac{H(X)}{n} + O(\frac{1}{n^2})$$

unde

$$H(X) = \mu_{2k}(X) - \mu_k^2(X) - 2k\mu_{k-1}(X)\mu_{k+1}(X) + k^2\mu_2(X)\mu_{k-1}^2(X)$$

şi $O(\frac{1}{n^2})$ este o funcţie asiptotică de ordinul lui $\frac{1}{n^2}$.

Demonstraţie. A se vedea Mihoc G., Firescu D. [9].

Teorema 4. Dacă colectivitatea admite momente centrate de ordinele 2, k - 1,k+1,2k , atunci statistica

$$S(X,n) = \frac{b_k(X,n) - \mu_k(X)}{\sqrt{\frac{H(X)}{n}}}$$

urmează asimptotic repartiţia N(0,1).

Demonstraţie. A se vedea Reischer Corina, Sămboan G., Theodorescu R.[15].

Observaţia 1. Valoarea lui $b_k(X,)$ pentru selecţia $(x_1,\ldots,x_n)$ fixată este

$$\bar{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$$

Observaţia 2. Intre momentele iniţiale şi centrate de selecţie există aceleaşi formule de legătură ca şi între momentele iniţiale şi centrate teoretice sau statistice. Aceste formule sînt:

$$b_k(X,n) = \sum_{j=0}^{k}(-1)^j C_k^j \, a_{k-j}(X,n)m^j(X,n)$$

$$a_k(X,n) = \sum_{j=0}^{k} C_k^j \, b_{k-j}(X,n)m^j(X,n)$$

pentru $k = 1,2,3,..$

§ 7. TEORIA SELECŢIEI DINTR-O COLECTIVITATE NORMALĂ UNIDIMENSIONALĂ.

Prin colectivitate normală unidimensională se înţelege o colectivitate la indivizii căreia, caracteristica urmărită X, are o repartiţie normală. In acest paragraf ne vom ocupa de repartiţiile probabilistice ale unor statistici în cazul cînd selecţiile se efectuează dintr-o colectivitate normală unidimensională.

Teorema 1. Dacă colectivitatea are repartiţia $N(m,\sigma^2)$, atunci media de selecţie urmează repartiţia $N(m,\frac{\sigma^2}{n})$.

Demonstraţie. Se foloseşte metoda funcţiei caracteristice. Conform teoremei 5, §3, cap.I, funcţia caracteristică a repartiţiei $N(m,\sigma^2)$ este

$$\varphi(t) = e^{mit - \frac{t^2\sigma^2}{2}}.$$

Ţinînd seama că variabilele aleatoare de selecţie X_i, $1 \leq i \leq n$, sînt independente şi au aceeaşi funcţie de repartiţie ca şi X, rezultă că, funcţia caracteristică a mediei de selecţie, va fi

$$\varphi(\cdot) = M\left[e^{itm(X,n)}\right] = M\left[e^{i\frac{t}{n}\sum_{k=1}^{n}X_k}\right] = \prod_{k=1}^{n} M\left[e^{i\frac{t}{n}X_k}\right] =$$

$$= \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{n}\right) = \prod_{k=1}^{n} \varphi_X\left(\frac{t}{n}\right) = \prod_{k=1}^{n} \left(e^{mi\frac{t}{n} - \frac{t^2\sigma^2}{2n^2}}\right) = e^{mit - \frac{t^2\sigma^2}{2n}}$$

care este tocmai funcţia caracteristică a repartiţiei $N(m, \frac{\sigma^2}{n})$.

Teorema 2. Dacă colectivitatea are repartiţia $N(m, \sigma^2)$, atunci statistica

$$U(X,n) = \frac{m(X,n) - m}{\frac{\sigma}{\sqrt{n}}}$$

urmează repartiţia $N(0,1)$.

Demonstraţie. Statistica $U(X,n)$ este tocmai variabila aleatoare normată a lui $m(X,n)$. Concluzia teoremei rezultă din teorema 1 a acestui paragraf şi teorema 6, §3, cap.I.

Observaţia 1. Conform teoremei 5, §3, cap.III, statistica $U(X,n)$ urmează asimptotic repartiţia $N(0,1)$ indiferent de repartiţia probabilistică a caracteristicii X. Prin urmare, ţinînd seama şi de teorema 2 a acestui paragraf se poate afirma că statistica $U(X,n)$ urmează repartiţia $N(0,1)$ atît pentru n fixat, cît şi pentru $n \to \infty$.

Teorema 3. Dacă colectivităţile C_j, $1 \leqslant j \leqslant k$, au repartiţiile $N(m_j, \sigma_j^2)$, $1 \leqslant j \leqslant k$, iar $m(X_j, n_j)$, $1 \leqslant j \leqslant k$, sînt mediile de selecţie ce corespund la k selecţii repetate, independente, efectuate din cele k colectivităţi şi $a_j \in R$, $1 \leqslant j \leqslant k$, atunci statistica

$$S(X_1, \ldots, X_k; n_1, \ldots, n_k) = \sum_{j=1}^{k} a_j m(X_j, n_j)$$

urmează repartiţia

$$N\left(\sum_{j=1}^{k} a_j m_j, \sum_{j=1}^{k} \frac{a_j^2 \sigma_j^2}{n_j}\right)$$

Demonstraţie. Conform teoremei 1 din acest paragraf, $m(X_j, n_j)$ urmează repartiţia $N(m_j, \frac{\sigma_j^2}{n_j})$, $1 \leqslant j \leqslant k$. Conform

teoremei 7, §3, cap.I variabila aleatoare $\sum_{j=1}^{k} a_j m(X_j,n_j)$ urmează repartiţia $N(\sum_{j=1}^{k} a_j m_j, \sum_{j=1}^{k} \frac{a_j^2 \sigma_j^2}{n_j})$.

In particular, dacă $a = 1$, $a_2 = -1$, $a_3 = a_4 = \ldots = a_k = 0$, atunci statistica

$$S(X_1,X_2;n_1,n_2) = m(X_1,n_1) - m(X_2,n_2)$$

urmează repartiţia $N(m_1 - m_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.

Teorema 4. Dacă colectivităţile C_i, $1 \leq i \leq 2$, au repartiţiile $N(m_i, \sigma_i^2)$ $1 \leq i \leq 2$, atunci statistica

$$U(X_1,X_2;n_1,n_2) = \frac{m(X_1,n_1) - m(X_2,n_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

urmează repartiţia $N(0,1)$.

Demonstraţie. Intrucît statistica $U(X_1,X_2;n_1,n_2)$ reprezintă variabila aleatoare normată a statisticii $m(X_1,n_1)-m(X_2,n_2)$, concluzia teoremei rezultă din particularizarea făcută în teorema 3 a acestui paragraf şi teorema 6, §3, cap.I.

Teorema 5. Dacă colectivitatea are repartiţia $N(0, \sigma^2)$, atunci statistica

$$n\, a_2(X,n) = \sum_{i=1}^{n} X_i^2$$

urmează repartiţia $\chi^2(n,\sigma)$.

Demonstraţie. Concluzia teoremei rezultă din independenţa variabilelor aleatoare de selecţie X_i, $1 \leq i \leq n$, şi teorema 10, §3, cap.I.

Teorema 6. Dacă colectivitatea are repartiţia $N(0, \sigma^2)$, atunci momentul iniţial de selecţie de ordinul doi $a_2(X,n)$ urmează o repartiţie a cărei densitate de repartiţie este

$$g(x) = \begin{cases} \dfrac{n^{\frac{n}{2}}\, x^{\frac{n}{2}-1}\, e^{-\frac{nx}{2\sigma^2}}}{2^{\frac{n}{2}}\, \sigma^n\, \Gamma(\frac{n}{2})} & x > 0 \\ \\ & x < 0 \end{cases}$$

Demonstraţie. Fie $Y = \sum_{i=1}^{n} X_i^2$, avînd densitatea de repartiţie $f(y)$ şi funcţia de repartiţie $F(y)$. Conform teoremei 5 din acest paragraf, avem

$$f(y) = \begin{cases} \dfrac{y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}}{2^{n/2}\sigma^n \Gamma(\frac{n}{2})} & y > 0 \\ 0 & y \leqslant 0 \end{cases}$$

Fie $G(x)$ şi $g(x)$, funcţia de repartiţie, respectiv, densitatea de repartiţie a lui $a_2(X,n)$. Intrucît $a_2(X,n) = \frac{Y}{n}$, avem

$$G(x) = P(\omega : a_2(X,n) < x) = P(\omega : \frac{Y}{n} < x) = P(\omega : Y < nx) =$$

$$= \begin{cases} F(nx) & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

Derivînd ultima relaţie în raport cu x, se obţine

$$G'(x) = g(x) = \begin{cases} nF'(nx) & x > 0 \\ 0 & x \leqslant 0 \end{cases} = \begin{cases} nf(nx) & x > 0 \\ 0 & x \leqslant 0 \end{cases} =$$

$$= \begin{cases} \dfrac{n^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{nx}{2\sigma^2}}}{2^{\frac{n}{2}}\sigma^n \Gamma(\frac{n}{2})} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

Teorema 7. Dacă colectivitatea are repartiţia $N(m, \sigma^2)$, atunci statistica

$$S(X,n) = \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)^2$$

urmează repartiţia $\chi^2(n,1)$.

Demonstraţie. Conform teoremei 6, §3, cap.I, variabilele aleatoare $\frac{X_i - m}{\sigma}$, $1 \leqslant i \leqslant n$, urmează repartiţia $N(0,1)$. Intrucît variabilele aleatoare $\frac{X_i - m}{\sigma}$, $1 \leqslant i \leqslant n$, sînt independente, concluzia teoremei rezultă din teorema 10, §3, cap.I.

Teorema 8. Dacă colectivitatea are repartiţia $N(m, \sigma^2)$, atunci

i) statistica $$S(X,n) = \frac{(n-1)d_*^2(X,n)}{\sigma^2}$$

urmează repartiţia $\chi^2(n-1,1)$.

ii) statisticile $m(X,n)$ şi $d_*^2(X,n)$ sînt independente

Demonstraţie. A se vedea Mihoc G., Firescu D. [9].

Observaţia 2. Ţinînd seama că

$$d_*^2(X,n) = \frac{n}{n-1}d^2(X,n)$$

rezultă

$$\frac{(n-1)d_*^2(X,n)}{\sigma^2} = \frac{nd^2(X,n)}{\sigma^2}.$$

Prin urmare, statistica $\frac{nd^2(X,n)}{\sigma^2}$ urmează repartiţia $\chi^2(n-1,1)$.

Teorema 9. Dacă colectivitatea are repartiţia $N(m,\sigma^2)$, atunci statistica

$$t(X,n) = \frac{m(X,n) - m}{\sqrt{\frac{d_*^2(X,n)}{n}}}$$

urmează repartiţia Student cu n-1 grade de libertate.

Demonstraţie. Conform teoremei 2 din acest paragraf, variabila aleatoare

$$U = \frac{m(X,n) - m}{\frac{\sigma}{\sqrt{n}}}$$

urmează repartiţia $N(0,1)$. Conform primei părţi a teoremei 8 din acest paragraf, variabila aleatoare

$$V = \frac{(n-1)d_*^2(X,n)}{\sigma^2}$$

urmează repartiţia $\chi^2(n-1,1)$. Conform părţii a doua a teoremei 8 din acest paragraf, variabilele aleatoare U şi V sînt independente. Aplicînd teorema 13, §3, cap.I, se obţine că statistica

$$t(X,n) = \frac{U}{\sqrt{\frac{V}{n-1}}} = \frac{m(X,n) - m}{\sqrt{d_*^2(X,n)}}$$

rmează repartiţia Student cu n-1 grade de libertate,

Observaţia 3. Ţinînd seama că

$$d^2_*(X,n) = \frac{n}{n-1} d^2(X,n)$$

tatistica t(X,n) se mai poate exprima prin

$$t(X,n) = \frac{m(X,n) - m}{\sqrt{\frac{d^2(X,n)}{n-1}}}$$

Teorema 10. Dacă colectivităţile C_i, $1 \leqslant i \leqslant 2$, au repartiiile $N(m_i, \sigma^2)$, $1 \leqslant i \leqslant 2$, şi $m(X_i,n_i)$, $d^2_*(X_i,n_i)$, $1 \leqslant i \leqslant 2$, înt mediile de selecţie, respectiv dispersiile de selecţie ce orespund la două selecţii independente, efectuate din cele dolă colectivităţi, atunci statistica

$$(X_1,X_2;n_1,n_2) = \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \; \frac{m(X_1,n_1)-m(X_2,n_2)-(m_1-m_2)}{\sqrt{(n_1-1)d^2_*(X_1,n_1)+(n_2-1)d^2_*(X_2,n_2)}}$$

rmează rep tiţia Student cu $n_1 + n_2 - 2$ grade de libertate.

Demonstraţie. Conform teoremei 4 a acestui paragraf, variabila aleatoare

$$U = \frac{m(X_1,n_1) - m(X_2,n_2) - (m_1 - m_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

urmează repartiţia N(0,1). Potrivit primei părţi a teoremei 8 din acest paragraf, statistica

$$\frac{(n_1-1)d^2_*(X_1,n_1)}{\sigma^2}$$

urmează repartiţia $\chi^2(n_1-1,1)$, iar statistica

$$\frac{(n_2-1)d^2_*(X_2,n_2)}{\sigma^2}$$

urmează repartiţia $\chi^2(n_2-1,1)$. Din independenţa selecţiilor rezultă independenţa statisticilor

$$\frac{(n_1-1)d_*^2(X_1,n_1)}{\sigma^2} \text{ şi } \frac{(n_2-1)d_*^2(X_2,n_2)}{\sigma^2} .$$

Conform teoremei 9, §3, cap.I, variabila aleatoare

$$V = \frac{(n_1-1)d_*^2(X_1,n_1) + (n_2-1)d_*^2(x_2,n_2)}{\sigma^2}$$

urmează repartiţia $\chi^2(n_1 + n_2 - 2,1)$. Aplicînd teorema 13, §3, cap.I, se obţine că statistica

$$t(X_1,X_2;n_1,n_2)=\frac{U}{\sqrt{\frac{V}{n_1+n_2-2}}}=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\,\frac{m(X_1,n_1-m(X_2,n_2)-(m_1-m_2)}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,m}}$$

urmează repartiţia Student cu n_1+n_2-2 grade de libertate.

Teorema 11. Dacă colectivităţile C_i, $1 \leqslant i \leqslant 2$, au repartiţiile $N(m_i,\sigma_i^2)$, $1 \leqslant i \leqslant 2$, şi $d_*^2(X_i,n_i)$, $1 \leqslant i \leqslant 2$, sînt dispersiile de selecţie ce corespund la două selecţii repetate, independente efectuate din cele două colectivităţi, atunci statistica

$$F(X_1,X_2;n_1,n_2) = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{d_*^2(X_1,n_1)}{d_*^2(X_2,n_2)}$$

urmează repartiţia Snedecor cu $n_1 - 1$ şi $n_2 - 1$ grade de libertate.

Demonstraţie. Conform primei părţi a teoremei 8 din acest paragraf, variabila aleatoare

$$U = \frac{(n_1-1)d_*^2(X_1,n_1)}{\sigma_1^2}$$

urmează repartiţia $\chi^2(n_1-1,1)$ şi variabila aleatoare

$$V = \frac{(n_2-1)d_*^2(x_2,n_2)}{\sigma_2^2}$$

urmează repartiţia $\chi^2(n_2-1,1)$. Din independenţa celor două selecţii rezultă independenţa variabilelor aleatoare U şi V. Aplicînd teorem 14, §3, cap.I, se obţine că, statistica

$$F(X_1,X_2,n_1,n_2) = \frac{n_2-1}{n_1-1}\,\frac{U}{V} = \frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{d_*^2(X_1,n_1)}{d_*^2(X_2,n_2)}$$

urmează repartiţia Snedecor cu n_1-1 şi n_2-1 grade de libertate.

Observaţia 5. Ţinînd seama că

$$d_*^2(X,n) = \frac{n}{n-1}\,d^2(X,n)$$

statistica $F(X_1,X_2;n_1,n_2)$, se mai poate exprima prin

$$F(X_1,X_2;n_1,n_2) = \frac{n_1(n_2-1)d^2(X_1,n_1)}{n_2(n_1-1)d^2(X_2,n_2)} .$$

P R O B L E M E

1. Să se arate că momentele centrate de ordinele trei şi patru ale mediei de selecţie sînt

$$\mu_3[m(X,n)] = \frac{\mu_3(X)}{n^2} \,,\quad \mu_4[m(X,n)] = \frac{3\mu_2^2(X)}{n^2} + \frac{\mu_4(X)-3\mu_2^2(X)}{n^3}$$

2. Repartiţia unei colectivităţi este dată de

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Să se arate că media de selecţie urmează aceeaşi repartiţie probabilistică.

3. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$X : \begin{bmatrix} 0 & 1 \\ q & p \end{bmatrix}, \quad p+q=1 .$$

Din colectivitate se efectuează selecţii repetate de volum $n = 3$. Să se determine repartiţia probabilistică, funcţia caracteristică, media şi dispersia mediei de selecţie.

4. Repartiţia unei colectivităţi este normală de medie $m = 0$ şi dispersie $\sigma^2 = 4$. Din colectivitate se efectuează

selecţii repetate de volum $n = 9$. Să se determine repartiţia probabilistică, funcţia caracteristică, media şi dispersia pentru media de selecţie şi momentul iniţial de selecţie de ordinul doi.

5. Caracteristica X a unei colectivităţi urmează o repartiţie normală de medie $m = -2$ şi dispersie $\sigma^2 = 3$. Din colectivitate se efectuează selecţii repetate de volum n. Să se determine repartiţia probabilistică a statisticii

$$S(X,16) = \sum_{i=1}^{16} Y_i^2$$

unde, pentru i fixat, Y_i este variabila aleatoare normată a variabilei aleatoare de selecţie X_i.

6. Repartiţia unei colectivităţi este normală de medie $m = 5$ şi dispersie $\sigma^2 = 9$. Din colectivitate se efectuează selecţii repetate de volum $n = 10$. Să se determine repartiţiile probabilistice ale dispersiilor de selecţie $d^2(X,10)$ şi $d_*^2(X,10)$.

Capitolul IV

TEORIA ESTIMAȚIEI

Scanned with OKEN Scanner

Pentru a putea rezolva orice problemă, cu privire la o colectivitate statistică, este necesar a cunoaşte funcţia de repartiţie $F(x)$ a colectivităţii.

Definiţia 1. Se spune că funcţia de repartiţie a unei colectivităţi este complet specificată dacă se cunoaşte forma funcţională a ei şi această formă nu conţine nici un parametru necunoscut.

Exemplu. Funcţia de repartiţie a unei colectivităţi Poisson de parametru $\lambda = 5$, dată de

$$F(x) = \sum_{k=o}^{[x]} e^{-5} \frac{5^k}{k!}$$

este o funcţie de repartiţie complet specificată.

Definiţia 2. Se spune că funcţia de repartiţie a unei colectivităţi este specificată dacă se cunoaşte forma funcţională a ei, dar această formă conţine cel puţin un parametru necunoscut.

Exemplu. Funcţia de repartiţie a unei colectivităţi, avînd repartiţia $N(m, \sigma^2)$, dată prin

$$F(x; m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-m)^2}{2\sigma^2}} dv$$

unde $m \in R$, $\sigma > 0$, este specificată, pentru că m şi σ^2 sînt parametri necunoscuţi.

Definiţia 3. Se numeşte spaţiul parametrilor mulţimea valorilor admisibile ale parametrilor ce intră într-o funcţie de repartiţie specificată.

Spaţiul parametrilor, pe care-l vom nota cu Λ , este un interval deschis al lui R (cazul unui singur parametru) sau un domeniu al spaţiului euclidian R^k (cazul a k parametri). O funcţie de repartiţie specificată o vom nota prin $F(x; \lambda)$, unde $\lambda \in \Lambda$, îl vom numi punct parametric.

Pentru a ajunge de la funcţia de repartiţie specificată la funcţia de repartiţie complet specificată se impune a estima parametrii necunoscuţi. In cele ce urmează ne vom ocupa de estimarea parametrilor unei funcţii de repartiţie specificată cu ajutorul funcţiilor de estimaţie şi al intervalelor de încredere. Estimarea parametrilor cu ajutorul funcţiilor de estimaţie poartă numele de estimare punctuală.

§ 1. FUNCŢIE DE ESTIMAŢIE CONSISTENTĂ.

Fie $F(x;\lambda)$, $\lambda \in \Lambda \subset R$, funcţia de repartiţie a unei colectivităţi.

Definiţia 1. Se spune că statistica $E(X,n)$ este o funcţie de estimaţie consistentă pentru parametrul λ, dacă şirul de statistici

$$(E(X,n))_{n \in N^*}$$

converge în probabilitate către λ.

Teorema 1. Dacă $E(X,n)$ este o funcţie de estimaţie consistentă pentru λ, iar a şi b sînt numere reale fixate, atunci statistica

$$\overline{E(X,n)} = \frac{n-a}{n-b} E(X,n)$$

este o funcţie de estimaţie consistentă pentru λ.

Demonstraţie. Din ipoteză, conform definiţiei 1, rezultă că, şirul de statistici $(E(X,n))_{n \in N^*}$ converge în probabilitate către λ. Pentru a arăta că şirul de statistici $\overline{(E(X,n))}_{n \in N^*}$ converge în probabilitate către λ, este suficient a arăta că, şirul de variabile aleatoare

$$\left(\frac{n-a}{n-b}\right)_{n \in N^*}$$

converge în probabilitate către 1. Avînd în vedere faptul că, convergenţa în medie pătratică implică convergenţa în probabilitate, vom arăta că

$$\left(\frac{n-a}{n-b}\right)_{n\in N^*}$$

converge în medie pătratică, către 1. Intr-adevăr, avem

$$\beta_2\left(\frac{n-a}{n-b}-1\right)=\beta_2\left(\frac{b-a}{n-b}\right)$$

Pentru n fixat, mărimea $\frac{b-a}{n-b}$, poate fi considerată ca o variabilă aleatoare ce ia o singură valoare, egală cu $\frac{b-a}{n-b}$, cu probabilitatea 1. Prin urmare,

$$\beta_2\left(\frac{b-a}{n-b}\right)=M\left[\left(\frac{b-a}{n-b}\right)^2\right]=\frac{(b-a)^2}{(n-b)^2}$$

Făcînd $n\to\infty$ în ultima relaţie, se obţine,

$$\lim_{n\to\infty}\beta_2\left(\frac{n-a}{n-b}-1\right)=\lim_{n\to\infty}\beta_2\left(\frac{b-a}{n-b}\right)=\lim_{n\to\infty}\frac{(b-a)^2}{(n-b)^2}=0\ .$$

Observaţie. Pe baza teoremei demonstrate, se poate afirma că, odată cu o funcţie de estimaţie consistentă, parametrul λ admite o infinitate de asemenea funcţii de estimaţie.

Definiţia 2. Prin estimaţie a parametrului λ se înţelege valoarea funcţiei de estimaţie consistentă pentru o selecţie $(x_1,x_2,\ldots,x_n)$ fixată.

§ 2. FUNCŢIE DE ESTIMAŢIE ABSOLUT CORECTĂ.

Fie $F(x;\lambda),\lambda\in\Lambda\subset R$, funcţia de repartiţie a unei colectivităţi.

Definiţia 1. Se numeşte funcţie de estimaţie absolut corectă pentru parametrul λ, o statistică $E(X,n)$, care îndeplineşte condiţiile:

a) $M\left[E(X,n)\right]=\lambda$

b) $\lim_{n\to\infty}D^2\left[E(X,n)\right]=0$

Teorema 1. O funcţie de estimaţie absolut corectă pentru parametrul λ este o funcţie de estimaţie consistentă.

Demonstraţie. Tinînd seama de definiţia 1, §1, cap.IV, trebuie să demonstrăm că $(E(X,n))_{n\in N^*}$ converge în probabilitate

către λ . Aceasta revine, conform definiţiei 14, §1, cap.I la a arăta că, pentru orice $\varepsilon>0$, are loc

$$\lim_{n\to\infty} P(|E(X,n) - \lambda| \geqslant \varepsilon) = 0$$

Intr-adevăr, aplicînd statisticii $E(X,n)$ inegalitatea lui Cebîşev se obţine

$$0 \leqslant P(|E(X,n) - M[E(X,n)]| \geqslant \varepsilon) \leqslant \frac{D^2[E(X,n)]}{\varepsilon^2}$$

pentru orice $\varepsilon>0$. Ţinînd seama de condiţia a) din definiţia 1, avem

$$0 \leqslant P(|E(X,n) - \lambda| \geqslant \varepsilon) \leqslant \frac{D^2[E(X,n)]}{\varepsilon^2}$$

Făcînd $n \to \infty$ în ultima relaţie, se obţine

$$0 \leqslant \lim_{n\to\infty} P(|E(X,n) - \lambda| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \lim_{n\to\infty} D^2[E(X,n)]$$

Ţinînd seama de condiţia b) din definiţia 1, rezultă că, pentru orice $\varepsilon>0$, are loc

$$\lim_{n\to\infty} P(|E(X,n) - \lambda| \geqslant \varepsilon) = 0$$

Observaţie. O funcţie de estimaţie absolut corectă se mai numeşte şi funcţie de estimaţie nedeplasată sau funcţie de estimaţie fără eroare sistematică.

Teorema 2. Momentul iniţial de selecţie de ordin k este funcţie de estimaţie absolut corectă pentru momentul iniţial de ordin k al colectivităţii.

Demonstraţie. Verificăm îndeplinirea condiţiilor a) şi b) din definiţia 1 dată în acest paragraf. Condiţia a) este îndeplinită datorită teoremei 1, §4, cap.III. Condiţia b) este îndeplinită datorită teoremei 2, §4, cap.III.

Pentru $k = 1$ teorema 2 afirmă că media de selecţie este funcţie de estimaţie absolut corectă pentru media colectivităţii.

Pe baza teoremei 2 se poate scrie

$$\alpha_k(X) \simeq \overline{\alpha}_k$$

Teorema 3. Dispersia de selecţie $d_*^2(X,n)$ este funcţie de estimaţie absolut corectă pentru dispersia colectivităţii.

Demonstrație. Verificăm îndeplinirea condițiilor a) și b) din definiția 1 dată în acest paragraf. Condiția a) este îndeplinită datorită teoremei 5, §5, cap.III. Condiția b) este îndeplinită datorită teoremei 6, §5, cap.III.

Pe baza teoremei 3 se poate scrie

$$D^2(X) \simeq s_*^2 \ .$$

Teorema 4. Pentru x fixat, funcția de repartiție de selecție este funcție de estimație absolut corectă pentru funcția de repartiție a colectivității.

Demonstrație. Condițiile a) și b) din definiția 1 dată în acest paragraf sînt îndeplinite, datorită teoremei 1,§2,cap.III.

Pe baza teoremei 4, pentru $\omega^{(n)}$ fixat, funcția de repartiție a colectivității este aproximată de funcția de frecvență cumulată (funcția de repartiție statistică).

§ 3. FUNCȚIE DE ESTIMAȚIE EFICIENTĂ .

Intrucît pentru un același parametru ce intră într-o funcție de repartiție specificată pot exista mai multe funcții de estimație absolut corectă, se pune problema alegerii „celei mai bune". De exemplu, dacă colectivitatea urmează o repartiție Poisson de parametru λ , atunci conform teoremei 5, §2,cap.I, media și dispersia colectivității sînt egale cu λ . Dacă λ se interpretează ca medie a colectivității, atunci conform teoremei 2, §2, cap.IV, parametrul λ este estimat absolut corect de către media de selecție. Dacă se interpretează ca dispersie a colectivității, atunci conform teoremei 3, §2, cap. IV, parametrul λ este estimat absolut corect de dispersia de selecție $d_*^2(X,n)$.

Fie $E(X,n)$ o funcţie de estimaţie absolut corectă pentru parametrul λ. Aplicînd statisticii $E(X,n)$ inegalitatea lui Cebîşev, urmează că, pentru orice $\varepsilon>0$, are loc

$$P(|E(X,n) - \lambda| \geqslant \varepsilon) \leqslant \frac{D^2[E(X,n)]}{\varepsilon^2} .$$

Ţinînd seama de inegalitatea lui Cebîşev, dintre mai multe funcţii de estimaţie absolut corectă pentru un acelaşi parametru λ, se consideră ca fiind „cea mai bună" aceea care are dispersia cea mai mică, întrucît ea realizează marginea superioară a probabilităţii $P(|E(X,n) - \lambda| \geqslant \varepsilon)$. Nu totdeauna există o funcţie de estimaţie absolut corectă de dispersie minimă. In cazul cînd există, aceasta se determină folosind

Teorema I. (Rao - Cramer)

I. Dacă $E(X,n)$ este o funcţie de estimaţie absolut corectă pentru parametrul $\lambda \in \Lambda \subset R$ din densitatea de repartiţie $f(x;\lambda)$, atunci are loc

$$D^2[E(X,n)] \geqslant \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2}$$

II. Funcţia de estimaţie absolut corectă $E(X,n)$ pentru care are loc

$$D^2[E(X,n)] = \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2}$$

este de forma

$$E(X,n) = \frac{1}{n}\sum_{i=1}^{n} L(X_i) .$$

III. Egalitatea

$$D^2[E(X,n)] = \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2}$$

are loc, dacă şi numai dacă

$$\ln f(x;\lambda) = A'(\lambda)[L(x) - \lambda] + A(\lambda) + N(x)$$

unde $L(x)$ şi $N(x)$ sînt funcţii numai de x, $A(\lambda)$ este o funcţie numai de λ, iar $A'(\lambda)$ este derivata lui $A(\lambda)$.

Demonstrație. Pentru demonstrație se fac, în plus, următoarele presupuneri:

i) derivata $\frac{\partial f(x;\lambda)}{\partial \lambda}$ există oricare ar fi $x \in R$.

ii) expresia $\int_{L^n} \prod_{i=1}^{n} f(x_i;\lambda)dx_i$ este derivabilă sub integrală în raport cu λ, iar L^n este spațiul valorilor de selecție

iii) oricare ar fi $\lambda \in \Lambda$, $M\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2 < +\infty$

iv) expresia $\int_{L^n} E(x_1,\ldots,x_n)\prod_{i=1}^{n} f(x_i;\lambda)dx_i$ este derivabilă sub integrală în raport cu λ.

I. Din faptul că $f(x;\lambda)$ este o densitate de repartiție urmează

$$\int_{-\infty}^{+\infty} f(x;\lambda)dx = 1 . \tag{1}$$

Derivînd (1) în raport cu λ și permutînd operația de derivare cu cea de integrare, lucru permis întrucît limitele de integrare nu depind de λ, avem

$$\int_{-\infty}^{+\infty} \frac{\partial f(x;\lambda)}{\partial \lambda} dx = 0 \tag{2}$$

ceea ce este echivalent cu

$$\int_{-\infty}^{+\infty} \frac{\partial \ln f(x;\lambda)}{\partial \lambda} f(x;\lambda)dx = 0 \tag{2'}$$

oricare ar fi $x \in R$. Dacă $(x_1,\ldots,x_n)$ este o selecție, atunci conform lui (2'), rezultă

$$\int_{-\infty}^{+\infty} \frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} f(x_i;\lambda)dx = 0 , \quad 1 \leqslant i \leqslant n \tag{3}$$

Relația (2') se mai exprimă prin

$$M\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right] = 0 \tag{2''}$$

iar relațiile (3) se exprimă prin

(3') $$M\left[\frac{\partial \ln f(X_i;\lambda)}{\partial \lambda}\right]=0\ ,\quad 1\leqslant i\leqslant n\ ,$$

unde X_i , $1\leqslant i\leqslant n$, sînt exemplarele lui X. Întrucît X_i, $1\leqslant i\leqslant n$, sînt independente, densitatea de repartiţie a vectorului aleator de selecţie $Y(\omega^{(n)}) = (X_1(\omega^{(n)}),\dots,X_n(\omega^{(n)}))$ este egală cu $\prod_{i=1}^{n} f(x_i;\lambda)$. Valoarea medie a statisticii $E(X,n)$ este

(4) $$M\left[E(X,n)\right] = \int_{L^n} E(x_1,\dots,x_n)\prod_{i=1}^{n} f(x_i;\lambda)dx_i\ .$$

Din faptul că $E(X,n)$ este o funcţie de estimaţie absolut corectă pentru λ , rezultă că $M\left[E(X,n)\right] = \lambda$. Prin urmare, relaţia (4) devine

(4') $$\int_{L^n} E(x_1,\dots,x_n)\prod_{i=1}^{n} f(x_i;\lambda)dx_i = \lambda\ .$$

Derivînd relaţia (4') în raport cu λ se obţine

(5) $$\int_{L^n} E(x_1,..,x_n)\sum_{i=1}^{n}\left[\frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} f(x_1;\lambda)\dots f(x_n;\lambda)\right]dx_1\dots dx_n=1$$

Întrucît $\prod_{i=1}^{n} f(x_i;\lambda)$ este densitatea de repartiţie a vectorului aleator de selecţie $Y(\omega^{(n)})$, rezultă

(6) $$\int_{L^n}\prod_{i=1}^{n} f(x_i;\lambda)dx_i = 1\ .$$

Derivînd (6) în raport cu λ se obţine

(7) $$\int_{L^n}\sum_{i=1}^{n}\left[\frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} f(x_1,\lambda)\dots f(x_n;\lambda)\right]dx_1\dots dx_n = 0\ .$$

Înmulţind (7) prin λ şi scăzînd rezultatul obţinut din (5), avem

(8) $$\int_{L^n}\left[E(x_1,\dots,x_n)-\lambda\right]\sum_{i=1}^{n}\left[\frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} f(x_1;\lambda)\dots f(x_n;\lambda)\right]dx_1\dots dx_n=1$$

ceea ce se mai exprimă prin

(8') $$M\left\{\left[E(X,n)-\lambda\right]\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial \lambda}\right]\right\}= 1$$

Aplicînd lui (8') inegalitatea lui Schwarz, §1, cap.I, se obţine

$$(9)\qquad M^2\left\{\left[E(X,n)-\lambda\right]\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]\right\}\leq$$

$$\leq M\left[E(X,n)-\lambda\right]^2 M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2$$

Ţinînd seama de (8'), inegalitatea (9) devine

$$(9')\qquad 1\leq M\left[E(X,n)-\lambda\right]^2 M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2$$

Întrucît independenţa variabilelor aleatoare de selecţie X_i, $1\leq i\leq n$, implică independenţa variabilelor aleatoare $\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}$, $1\leq i\leq n$, avem

$$M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 = M\left\{\sum_{i=1}^{n}\left[\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 +\right.$$

$$(10)\qquad \left. +2\sum_{j\neq k}\frac{\partial \ln f(X_j;\lambda)}{\partial\lambda}\,\frac{\partial \ln f(X_k;\lambda)}{\partial\lambda}\right\} = \sum_{i=1}^{n}M\left[\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 +$$

$$+\,2\sum_{j\neq k}M\left[\frac{\partial \ln f(X_j;\lambda)}{\partial\lambda}\right]M\left[\frac{\partial \ln f(X_k;\lambda)}{\partial\lambda}\right].$$

Folosind relaţiile (3'), relaţia (10) devine

$$(11)\qquad M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 = \sum_{i=1}^{n}M\left[\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2$$

Întrucît X_i, $1\leq i\leq n$, au aceeaşi funcţie de repartiţie ca şi X, avem

$$(12)\qquad M\left[\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 = M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2,\quad 1\leq i\leq n$$

Ţinînd seama de (12), relaţia (11) devine

$$(13)\qquad M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2 = n\,M\left[\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2$$

Avînd în vedere (13) şi faptul că

$$M\left[E(X,n)-\lambda\right]^2 = D^2\left[E(X,n)\right]$$

relaţia (9') devine

$$1 \leqslant D^2\left[E(X,n)\right] n\, M\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2$$

de unde

$$D^2\left[E(X,n)\right] \geqslant \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2}$$

Deci, prima parte a teoremei este demonstrată.

II. Inegalitatea (9') se transformă în egalitate, dacă şi numai dacă, există o funcţie $K(\lambda)$, aceeaşi pentru orice selecţie $(x_1,\ldots,x_n)$, astfel încît să aibă loc

$$(14) \qquad \sum_{i=1}^{n} \frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} = K(\lambda)\left[E(x_1,\ldots,x_n) - \lambda\right] -$$

Pentru selecţia $x_1 = x$, $x_2 = x_3 = \ldots = x_n = 0$, relaţia (14) devine

$$(15) \qquad \frac{\partial \ln f(x;\lambda)}{\partial \lambda} = K(\lambda)\left[E(x) - \lambda\right]$$

sau

$$(15') \qquad \frac{\partial \ln f(x;\lambda)}{\partial \lambda} = F(\lambda)Q(x) + G(\lambda)$$

Scriind relaţia (15') pentru valorile de selecţie x_i, $1 \leqslant i \leqslant n$, avem

$$(16) \qquad \frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} = F(\lambda)Q(x_i) + G(\lambda), \quad 1 \leqslant i \leqslant n$$

Din relaţia (14), rezultă

$$(17) \qquad E(x_1,\ldots,x_n) = \frac{1}{K(\lambda)} \sum_{i=1}^{n} \frac{\partial \ln f(x_i;\lambda)}{\partial \lambda} + \lambda$$

Ţinînd seama de (16), relaţia (17) devine

$$(17') \qquad E(x_1,\ldots,x_n) = \frac{1}{K(\lambda)} \sum_{i=1}^{n} \left[F(\lambda)Q(x_i) + G(\lambda)\right] + \lambda$$

sau

$$(17'') \qquad E(x_1,\ldots,x_n) = \frac{F(\lambda)}{K(\lambda)} \sum_{i=1}^{n} Q(x_i) + \frac{nG(\lambda)}{K(\lambda)} + \lambda \ .$$

Notăm

$$(18) \qquad \frac{F(\lambda)}{K(\lambda)} = h \ , \quad \frac{nG(\lambda)}{K(\lambda)} + \lambda = g \ .$$

Considerăm funcţia

$$(19) \qquad L(x) = nhQ(x) + g \ .$$

Ținînd seama de (18) și (19), relația (17") devine

$$(20)\quad E(x_1,\dots,x_n) = h\sum_{i=1}^{n}\frac{L(x_i)-g}{nh} + g = \frac{1}{n}\sum_{i=1}^{n}L(x_i)$$

Inlocuind în (20), valorile de selecție x_i, $1\leqslant i\leqslant n$, respectiv cu variabilele aleatoare de selecție X_i, $1\leqslant i\leqslant n$, se obține

$$(21)\qquad E(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^{n}L(X_i)$$

Prin urmare, partea a doua a teoremei este demonstrată .

III. Vom arăta că dacă

$$D^2\left[E(X,n)\right] = \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2}$$

ceea ce este echivalent cu

$$(22)\quad 1 = M\left[E(X,n)-\lambda\right]^2 M\left[\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}\right]^2$$

atunci

$$\ln f(x;\lambda) = A'(\lambda)\left[L(x)-\lambda\right] + A(\lambda) + N(x)$$

Exprimînd (15') în funcție de $K(\lambda)$ și $L(x)$, avem

$$(23)\quad \frac{\partial \ln f(x;\lambda)}{\partial\lambda} = hK(\lambda)\frac{L(x)-g}{nh} + \frac{g-\lambda}{n}K(\lambda) =$$

$$= \frac{K(\lambda)}{n}\left[L(x)-\lambda\right].$$

Integrînd (23) în raport cu λ , se obține

$$\ln f(x;\lambda) = A'(\lambda)\left[L(x)-\lambda\right] + A(\lambda) + N(x) .$$

Arătăm acum că dacă

$$(24)\quad \ln f(x;\lambda) = A'(\lambda)\left[L(x)-\lambda\right] + A(\lambda) + N(x)$$

ceea ce este echivalent cu

$$(24')\qquad f(x;\lambda) = e^{A'(\lambda)\left[L(x)-\lambda\right]+A(\lambda)+N(x)}$$

atunci

$$D^2\left[E(X,n)\right] = \frac{1}{nM\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2}$$

unde $E(X,n)$ este dat de (21).

Din faptul că (24') este o densitate de repartiție, rezultă

$$(25)\qquad \int_{-\infty}^{+\infty} e^{A'(\lambda)[L(x)-\lambda]+A(\lambda)+N(x)}\,dx = 1$$

Derivînd (25) în raport cu λ se obţine

$$(26)\qquad \int_{-\infty}^{+\infty} e^{A'(\lambda)[L(x)-\lambda]+A(\lambda)+N(x)}\,A''(\lambda)[L(x)-\lambda]\,dx = 0$$

sau ţinînd seama de (24),

$$(26')\qquad A''(\lambda)\int_{-\infty}^{+\infty}[L(x)-\lambda]f(x;\lambda)dx = 0$$

In ipoteza $A''(\lambda) \neq 0$, din (26') rezultă

$$(27)\qquad \int_{-\infty}^{+\infty}[L(x)-\lambda]f(x;\lambda)dx = 0$$

ceea ce se mai poate exprima prin

$$(27')\qquad M[L(X)-\lambda] = 0$$

sau

$$(27'')\qquad M[L(X)] = \lambda$$

Ţinînd seama că X_i, $1 \leq i \leq n$, au aceeaşi funcţie de repartiţie ca şi X, rezultă

$$(27'')\qquad M[L(X_i)] = \lambda\,,\quad 1 \leq i \leq n$$

Derivînd (26') în raport cu λ şi ţinînd seama de (27), se obţine

$$(28)\qquad A''(\lambda)\left\{-1 + A''(\lambda)\int_{-\infty}^{+\infty}[L(x)-\lambda]^2 f(x;\lambda)dx\right\} = 0$$

In presupunerea $A''(\lambda) \neq 0$, din (28) rezultă

$$(29)\qquad \int_{-\infty}^{+\infty}[L(x)-\lambda]^2 f(x;\lambda)dx = \frac{1}{A''(\lambda)}$$

care, pe baza lui (27''), se mai poate scrie

$$(29')\qquad \int_{-\infty}^{+\infty}[L(x)-M(L(X))]^2 f(x;\lambda)dx = \frac{1}{A''(\lambda)}$$

adică

$$(29'')\qquad D^2[L(X)] = \frac{1}{A''(\lambda)}$$

Ţinînd seama de (29'') şi faptul că X_i, $1 \leq i \leq n$, sînt

independente şi au aceeaşi funcţie de repartiţie ca şi X, avem

$$(30)\qquad D^2\left[E(X,n)\right] = D^2\left[\frac{1}{n}\sum_{i=1}^{n}L(X_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n}D^2\left[L(X_i)\right] =$$

$$= \frac{D^2[L(X)]}{n} = \frac{1}{nA''(\lambda)}$$

Vom arăta că

$$(31)\qquad M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2 = A''(\lambda)$$

Din (24) rezultă

$$(32)\qquad \frac{\partial \ln f(x;\lambda)}{\partial\lambda} = A''(\lambda)\left[L(x) - \lambda\right]$$

Ţinînd seama de (29") şi (32), avem

$$M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2 = \int_{-\infty}^{+\infty}\left[\frac{\partial \ln f(x;\lambda)}{\partial\lambda}\right]^2 f(x;\lambda)dx =$$

$$= A''^2(\lambda)\int_{-\infty}^{+\infty}\left[L(x) - \lambda\right]^2 f(x;\lambda)dx = A''^2(\lambda)D^2[L(X)] = A''(\lambda)$$

ceea ce reprezintă tocmai relaţia (31).

<u>Observaţia 1</u>. Are loc relaţia

$$M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2 = -M\left[\frac{\partial^2 \ln f(X;\lambda)}{\partial\lambda^2}\right]$$

Într-adevăr, pe baza relaţiilor (2") şi (2'), rezultă

$$(33)\qquad M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right] = \int_{-\infty}^{+\infty}\frac{\partial \ln f(x;\lambda)}{\partial\lambda} f(x;\lambda)dx = 0$$

Avem

$$(34)\qquad M\left[\frac{\partial \ln f(X;\lambda)}{\partial\lambda}\right]^2 = \int_{-\infty}^{+\infty}\left[\frac{\partial \ln f(x;\lambda)}{\partial\lambda}\right]^2 f(x;\lambda)dx =$$

$$= \int_{-\infty}^{+\infty}\left[\frac{1}{f(x;\lambda)}\frac{\partial f(x;\lambda)}{\partial\lambda}\right]^2 f(x;\lambda)dx = \int_{-\infty}^{+\infty}\left[\frac{\partial f(x;\lambda)}{\partial\lambda}\right]^2 \frac{1}{f(x;\lambda)}dx$$

Derivînd (33) în raport cu λ, se obţine

$$(35)\qquad \int_{-\infty}^{+\infty}\frac{\partial^2 \ln f(x;\lambda)}{\partial\lambda^2} f(x;\lambda)dx + \int_{-\infty}^{+\infty}\frac{\partial \ln f(x;\lambda)}{\partial\lambda}\frac{\partial f(x;\lambda)}{\partial\lambda}dx = 0$$

ceea ce este echivalent cu

$$5') \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2} f(x;\lambda)dx + \int_{-\infty}^{+\infty} \left[\frac{\partial f(x;\lambda)}{\partial \lambda}\right]^2 \frac{1}{f(x;\lambda)} dx = 0$$

ținînd seama de (34) relația (35') se mai scrie sub forma

$$5'') \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2} f(x;\lambda)dx + M\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2 = 0$$

u

$$5''') \quad M\left[\frac{\partial^2 \ln f(X;\lambda)}{\partial \lambda^2}\right] = -M\left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2$$

ea ce constituie relația de demonstrat.

Observația 2. Teorema Rao-Cramer se poate demonstra și în zul cînd caracteristica X are o repartiție probabilistică discretă, în condiții anumite impuse probabilității

$$P(X;\lambda) = P(\omega : X(\omega) = a_i) , \quad i \in I ,$$

de I este o mulțime cel mult numărabilă.

Observația 3. Dacă se notează

$$(\lambda) = \begin{cases} \left[\frac{\partial \ln f(X;\lambda)}{\partial \lambda}\right]^2 = -M\left[\frac{\partial^2 \ln f(X;\lambda)}{\partial \lambda^2}\right] \text{ pentru X continuă} \\ M\left[\frac{\partial \ln P(X;\lambda)}{\partial \lambda}\right]^2 = -M\left[\frac{\partial^2 \ln P(X;\lambda)}{\partial \lambda^2}\right] \text{ pentru X discretă} \end{cases}$$

tunci, inegalitatea din teorema Rao - Cramer, se scrie

$$D^2\left[E(X,n)\right] \geqslant \frac{1}{nR(\lambda)} .$$

Definiția 1. Se numește funcție de estimație eficientă pentru arametrul λ, o funcție de estimație absolut corectă $E(X,n)$, entru care are loc

$$D^2\left[E(X,n)\right] = \frac{1}{nR(\lambda)}$$

otrivit definiției 1, pentru o funcție de estimație eficientă, egalitatea din teorema Rao-Cramer, devine egalitate.

Definiția 2. Se numește eficiență a funcției de estimație bsolut corectă $E(X,n)$ mărimea

$$e\left[E(X,n)\right] = \frac{\frac{1}{nR(\lambda)}}{D^2\left[E(X,n)\right]} .$$

Observaţia 4. Din definiţia 2 şi inegalitatea din teorema Rao - Cramer, rezultă

$$0 \leqslant e\left[E(X,n)\right] \leqslant 1$$

oricare ar fi funcţia de estimaţie absolut corectă $E(X,n)$.

Avînd în vedere definiţiile 1 şi 2, rezultă

Definiţia 3. Se numeşte funcţie de estimaţie eficientă o funcţie de estimaţie absolut corectă pentru care eficienţa este egală cu 1 .

Definiţia 4. Se numeşte funcţie de estimaţie asimptotic eficientă o funcţie de estimaţie absolut corectă a cărei eficienţă satisface relaţia

$$\lim_{n\to\infty} e\left[E(X,n)\right] = 1$$

Teorema 2. Două funcţii de estimaţie eficientă sînt egale aproape sigur.

Demonstraţie. Fie $E_1(X,n)$ şi $E_2(X,n)$ funcţii de estimaţie eficientă pentru parametrul λ . Rezultă că au loc relaţiile

$$M\left[E_1(X,n)\right] = M\left[E_2(X,n)\right] = \lambda$$

$$\lim_{n\to\infty} D^2\left[E_1(X,n)\right] = \lim_{n\to\infty} D^2\left[E_2(X,n)\right] = 0$$

$$D^2\left[E_1(X,n)\right] = D^2\left[E_2(X,n)\right] = V^2$$

unde

$$V^2 = \min_{E(X,n)} D^2\left[E(X,n)\right]$$

iar $E(X,n)$ este o funcţie de estimaţie absolut corectă pentru λ . Fie k o constantă oarecare reală. Atunci, statistica

$$E(X,n) = kE_1(X,n) + (1-k)E_2(X,n)$$

este o funcţie de estimaţie absolut corectă pentru λ . Într-adevăr, ţinînd seama de relaţiile de mai înainte, avem

$$M\left[E(X,n)\right] = kM\left[E_1(X,n)\right] + (1-k)M\left[E_2(X,n)\right] = \lambda$$

Calculînd dispersia statisticii $E(X,n)$, se obţine

$$D^2\left[E(X,n)\right] = k^2D^2\left[E_1(X,n)\right] + (1-k)^2D^2\left[E_2(X,n)\right] + 2k(1-k)\operatorname{cov}.(E_1,E_2) =$$

$$= k^2V^2 + (1-k)^2V^2 + 2k(1-k)\,\rho(E_1,E_2) = V^2\left[2(1-\rho)k^2 - 2(1-\rho)k + 1\right]$$

Intrucît $E(X,n)$ este numai o funcţie de estimaţie absolut corectă, urmează că

$$D^2[E(X,n)] \geqslant V^2$$

adică

$$V^2[2(1-\rho)k^2 - 2(1-\rho)k + 1] \geqslant V^2$$

sau

$$2(1-\rho)k^2 - 2(1-\rho)k \geqslant 0$$

Pentru ca ultima relaţie să aibă loc oricare ar fi $k \in R$, trebuie îndeplinite condiţiile

$$(1-\rho)^2 \leqslant 0 \ , \ 1-\rho \geqslant 0$$

de unde rezultă că $\rho = 1$. Din $\rho(E_1,E_2) = 1$ rezultă $cov.(E_1,E_2) = V^2$. Pentru a demonstra că $E_1(X,n)$ şi $E_2(X,n)$ sînt egale aproape sigur este suficient a arăta că ,

$$(E_1(X,n) - E_2(X,n))_{n \in N^*}$$

converge în medie pătratică către zero. Intr-adevăr, pentru n fixat, avem

$$M[E_1(X,n) - E_2(X,n)]^2 = M[(E_1(X,n) - \lambda) - (E_2(X,n) - \lambda)]^2 =$$

$$= M[E_1(X,n) - \lambda]^2 + M[E_2(X,n) - \lambda]^2 -$$

$$- 2M\{[E_1(X,n) - \lambda][E_2(X,n) - \lambda]\} =$$

$$= D^2[E_1(X,n)] + D^2[E_2(X,n)] - 2\,cov.(E_1,E_2) = 0$$

Atunci

$$\lim_{n \to \infty} M[E_1(X,n) - E_2(X,n)]^2 = 0$$

şi teorema este demonstrată.

Observaţia 5. Din faptul că, coeficientul de corelaţie a funcţiilor de estimaţie eficientă $E_1(X,n)$ şi $E_2(X,n)$ este egal cu 1 , rezultă

$$E_2(X,n) = aE_1(X,n) + b$$

unde $a,b \in R$, şi cum

$$D^2[E_2(X,n)] = D^2[E_1(X,n)]$$

se obţine

$$E_2(X,n) = E_1(X,n) + b$$

§ 4. FUNCŢIE DE ESTIMAŢIE CORECTĂ

Fie $F(x;\lambda)$, $\lambda \in \Lambda \subset R$, funcţia de repartiţie a unei colectivităţi.

Definiţia 1. Se numeşte funcţie de estimaţie corectă pentru parametrul λ, o statistică $E(X,n)$ pentru care au loc condiţiile:

$$a) \quad M[E(X,n)] = \lambda + \alpha(\lambda,n)$$

unde $\lim_{n\to\infty} \alpha(\lambda,n) = 0$

$$b) \quad \lim_{n\to\infty} D^2[E(X,n)] = 0$$

Teorema 1. O funcţie de estimaţie corectă pentru parametrul λ este o funcţie de estimaţie consistentă pentru λ.

Demonstraţie. A se vedea demonstraţia teoremei 1, §2, cap. IV.

Observaţia 1. O funcţie de estimaţie corectă se mai numeşte funcţie de estimaţie deplasată sau funcţie de estimaţie cu eroare sistematică. Termenul $\alpha(\lambda,n)$, pentru n fixat, se numeşte eroare sistematică. In cazul unei funcţii de estimaţie absolut corectă eroarea sistematică este nulă.

Teorema 2. Dispersia de selecţie $d^2(X,n)$ este funcţie de estimaţie corectă pentru dispersia colectivităţii.

Demonstraţie. Verificăm îndeplinirea condiţiilor a) şi b) din definiţia 1 dată în acest paragraf. Condiţia a) este îndeplinită datorită teoremei 3, §5, cap.III. Condiţia b) este îndeplinită datorită teoremei 4, §5, cap.III.

Pe baza teoremei 2 se poate scrie

$$D^2(X) \simeq s^2 .$$

Teorema 3. Momentul centrat de selecţie de ordin k este funcţie de estimaţie corectă pentru momentul centrat de ordin k al colectivităţii.

Demonstrație. Verificăm îndeplinirea condițiilor a) și b) din definiția 1 dată în acest paragraf. Condiția a) este îndeplinită datorită teoremei 2, §6, cap.III. Condiția b) este îndeplinită datorită teoremei 3, §6, cap.III.

Observația 2. In anumite condiții, se poate demonstra că, dacă $E(X,n)$ este o funcție de estimație corectă pentru parametrul λ, atunci are loc inegalitatea

$$M[E(X,n) - \lambda]^2 \geqslant \frac{[1 + \alpha'(\lambda,n)]^2}{nR(\lambda)}$$

unde $\alpha'(\lambda,n)$ este derivata lui $\alpha(\lambda,n)$ în raport cu λ, iar $R(\lambda)$ are semnificația din observația 3, §3, cap.IV. In cazul cînd $E(X,n)$ este o funcție de estimație absolut corectă, $\lambda = M[E(X,n)]$, $\alpha'(\lambda,n) = 0$ și inegalitatea de mai înainte devine inegalitatea din teorema 1, §3, cap.IV.

§ 5. FUNCȚIE DE ESTIMAȚIE SUFICIENTĂ

Fie $f(x;\lambda), \lambda \in \Lambda \subset R$, densitatea de repartiție a unei colectivități la indivizii căreia se urmărește caracteristica X și L^n spațiul valorilor de selecție repetată de volum n. Considerăm aplicația

$$V : L^n \times \Lambda \longrightarrow R$$

definită prin

$$V(x_1,\ldots,x_n;\lambda) = \prod_{k=1}^{n} f(x_k;\lambda)$$

Dacă λ este fixat, atunci $V(x_1,\ldots,x_n;\lambda)$ reprezintă densitatea de repartiție a vectorului de selecție repetată de volum n, $Y(\omega^{(n)})$ al caracteristicii X. Dacă valorile de selecție $(x_1,\ldots,x_n)$ sînt fixate, atunci $V(x_1,\ldots,x_n;\lambda)$ este o funcție de parametrul λ numită verosimilitate.

Observație. Dacă repartiția colectivității este de tip discret, atunci aplicația V se definește prin

$$V(x_1,\dots,x_n;\lambda) = \prod_{k=1}^{n} p(x_k;\lambda)$$

unde $p(x;\lambda) = P(X = x)$.

Definiţia 1. Se spune că statistica $S(X,n)$ este o statistică suficientă pentru familia de densităţi de repartiţie $f(x;\lambda)$, dacă are loc

$$V(x_1,\dots,x_n;\lambda) = g[S(x_1,\dots,x_n),\lambda]\, h(x_1,\dots,x_n)$$

unde $g[S(x_1,\dots,x_n),\lambda]$ este densitatea de repartiţie a statisticii $S(X,n)$.

Definiţia 2. Se numeşte funcţie de estimaţie suficientă pentru parametrul λ , o funcţie de estimaţie absolut corectă care este o statistică suficientă pentru familia de densităţi de repartiţie $f(x;\lambda)$.

Teorema 1. O condiţie necesară şi suficientă pentru ca familia de densităţi de repartiţie $f(x;\lambda)$ să admită o statistică suficientă este ca

$$\ln f(x;\lambda) = A_1(\lambda)\Psi_1(x) + A_0(\lambda) + \Psi_0(x)$$

Orice funcţie de estimaţie suficientă a parametrului λ este o funcţie de $\sum_{i=1}^{n} \Psi_1(X_i)$.

Demonstraţie. A se vedea Onicescu O., Mihoc Gh. [12] .

Teorema 2. Orice funcţie de estimaţie eficientă este şi funcţie de estimaţie suficientă.

Demonstraţie. Fie $S(X,n)$ o funcţie de estimaţie eficientă pentru parametrul λ din familia de densităţi de repartiţie $f(x;\lambda)$. Conform teoremei 1, §3, cap.IV, rezultă

$$f(x;\lambda) = e^{A'(\lambda)[L(x)-\lambda]+A(\lambda)+N(x)}$$

iar funcţia de estimaţie eficientă este de forma

$$S(X,n) = \frac{1}{n} \sum_{i=1}^{n} L(X_i)$$

Densitatea de repartiţie a vectorului de selecţie $Y(\omega^{(n)})$ este

$$V(x_1,\dots,x_n;\lambda) = f(x_1;\lambda)\dots f(x_n;\lambda) =$$

$$= e^{A'(\lambda)[L(x_1)-\lambda]+A(\lambda)+N(x_1)} \dots e^{A'(\lambda)[L(x_n)-\lambda]+A(\lambda)+N(x_n)}$$

$$= e^{A'(\lambda)[nS(x_1,\dots,x_n)-n\lambda]+nA(\lambda)} \cdot e^{\sum_{i=1}^{n} N(x_i)}$$

ceea ce, conform definiţiei 1, înseamnă că $S(X,n)$ este o statistică suficientă pentru familia de densităţi de repartiţie $f(x;\lambda)$.

§ 6. METODE DE ESTIMARE PUNCTUALĂ A PARAMETRILOR.

In cele ce urmează vom prezenta următoarele metode de estimare punctuală a parametrilor unei legi de probabilitate: metoda momentelor, metoda verosimilităţii maxime, metoda minimului hi-pătrat şi metoda modificată a minimului lui hi-pătrat.

Metoda momentelor. Această metodă a fost propusă de către K. Pearson. Fie $F(x;\lambda_1,\dots,\lambda_k)$ funcţia de repartiţie a unei colectivităţi la indivizii căreia se urmăreşte caracteristica X. Presupunem că există şi sînt finite momentele iniţiale ale colectivităţii de ordinele de care este nevoie în această metodă. Intrucît funcţia de repartiţie a colectivităţii este specificată, se pot calcula momentele iniţiale ale colectivităţii şi aceste momente vor fi funcţii de parametrii $\lambda_1,\dots,\lambda_k$. Notăm cu $\alpha_r(X;\lambda_1,\dots,\lambda_k)$ momentul iniţial de ordin r al colectivităţii. Din colectivitate se efectuează o selecţie repetată de volum n. Pe baza acestei selecţii trebuie să determinăm estimaţii pentru parametri necunoscuţi $\lambda_1,\lambda_2,\dots,\lambda_k$.

Metoda momentelor constă în egalarea a k momente iniţiale ale colectivităţii cu momentele iniţiale statistice de acelaşi ordin. Sistemul de relaţii astfel obţinut se consideră ca un sistem de

ecuaţii în necunoscutele $\lambda_1,\dots,\lambda_k$. Egalînd primele k momente iniţiale se obţine sistemul de ecuaţii

$$\alpha_r(X;\lambda_1,\dots,\lambda_k)=\bar{\alpha}_r \quad , \quad 1\leq r\leq k .$$

Dacă acest sistem are soluţie unică, atunci prin rezolvarea lui se obţin estimaţiile $\bar{\lambda}_i$, $1\leq i\leq k$.

Observaţia 1. O justificare a metodei momentelor o constituie teorema 2, §2, cap.IV, conform căreia puteam scrie

$$\alpha_r(X;\lambda_1,\dots,\lambda_k)\simeq\bar{\alpha}_r$$

Observaţia 2. Metoda momentelor nu se poate aplica pentru estimarea punctuală a parametrilor unei repartiţii specificate ce nu admite momente iniţiale de nici un ordin. Un exemplu de acest fel îl constituie repartiţia Cauchy a cărei densitate de repartiţie este

$$f(x;\mu)=\frac{1}{\pi\left[1+(x-\mu)^2\right]}$$

Observaţia 3. Prin metoda momentelor se obţin funcţii de estimaţie absolut corectă. Estimaţia $\bar{\lambda}_i$ este valoarea funcţie de estimaţie $S_i(X,n)$ pentru selecţia considerată.

Metoda verosimilităţii maxime. Fie $V(x_1,\dots,x_n;\lambda)$ aplicaţia definită în §5, cap.IV.

Definiţia 1. Se spune că estimaţia $S(x_1,\dots,x_n)$ a parametrului λ este o estimaţie de verosimilitate maximă dacă ea reprezintă un punct de maxim pentru verosimilitate.

Presupunem că oricare ar fi $(x_1,x_2,\dots,x_n)\in L^n$ şi orice $\lambda\in\Lambda\subset R$, are loc $V(x_1,\dots,x_n;\lambda)>0$

Pe baza acestei presupuneri are sens a considera $\ln V(x_1,\dots,x_n;\lambda)$. Evident, funcţiile $V(x_1,\dots,x_n;\lambda)$ şi $\ln V(x_1,\dots,x_n;\lambda)$ au aceleaşi puncte de extrem. Din definiţia 1 dată în acest paragraf, rezultă că estimaţia de verosimilitate maximă $S(x_1,\dots,x_n)$ este o soluţie a ecuaţiei

$$\frac{\partial V(x_1,\dots,x_n;\lambda)}{\partial\lambda} = 0 \tag{1}$$

sau a ecuaţiei echivalente

$$\frac{\partial \ln V(x_1,\dots,x_n;\lambda)}{\partial\lambda} = 0 \tag{2}$$

Ecuaţia (2) poartă numele de <u>ecuaţia de verosimilitate</u> şi ea reprezintă condiţia necesară de extrem pentru verosimilitatea $V(x_1,\dots,x_n;\lambda)$. Estimaţiile de verosimilitate maximă se află printre soluţiile ecuaţiei de verosimilitate. Dintre soluţiile ecuaţiei de verosimilitate, numai cele ce depind de valorile de selecţie şi pentru care are loc

$$\frac{\partial^2 \ln V(x_1,\dots,x_n;\lambda)}{\partial\lambda^2} < 0$$

vor fi estimaţii de verosimilitate maximă. In cazul cînd verosimilitatea adm mai multe puncte de maxim, drept estimaţie de verosimilit aximă se va considera maximul absolut al verosimilităţii. Dacă $S(x_1,x_2,\dots,x_n)$ este o estimaţie de verosimilitate maximă, atunci statistica $S(X_1,X_2,\dots,X_n)$ se numeşte funcţie de estimaţie de verosimilitate maximă.

In cazul cînd funcţia de repartiţie a colectivităţii depinde de parametrii $\lambda_1, \lambda_2,\dots,\lambda_k$, atunci condiţia de extrem pentru verosimilitatea $V(x_1,\dots,x_n;\lambda_1,\lambda_2,\dots,\lambda_k)$ conduce la un sistem de ecuaţii de verosimilitate maximă, de forma

$$\frac{\partial \ln V(x_1,\dots,x_n;\lambda_1,\lambda_2,\dots,\lambda_k)}{\partial\lambda_i} = 0, \quad 1 \leq i \leq k \tag{3}$$

O condiţie suficientă pentru ca o soluţie $\lambda_1^o,\dots,\lambda_k^o$ a sistemului de ecuaţii (3) să reprezinte estimaţii de verosimilitate maximă pentru parametrii $\lambda_1,\lambda_2,\dots,\lambda_k$ este ca matricea

$$\left\| \frac{\partial^2 \ln V(x_1,\dots,x_n;\lambda_1,\dots,\lambda_k)}{\partial\lambda_r\,\partial\lambda_s} \right\|_{(\lambda_1^o,\dots,\lambda_k^o)}$$

să fie negativ definită.

Teorema 1. Orice funcţie de estimaţie eficientă verifică ecuaţiade verosimilitate.

Demonstraţie. Fie $E(X,n)$ o funcţie de estimaţie eficientă pentru parametrul λ din densitatea de repartiţie $f(x;\lambda)$. Conform teoremei 1,§3, cap.IV, rezultă

$$\ln f(x;\lambda) = A'(\lambda)[L(x) - \lambda] + A(\lambda) + N(x)$$

şi

$$E(X,n) = \frac{1}{n}\sum_{i=1}^{n} L(X_i)$$

In acest caz, ecuaţia de verosimilitate (2) devine

$$\sum_{k=1}^{n} \frac{\partial \ln f(x_k;\lambda)}{\partial \lambda} = 0$$

sau

$$A''(\lambda)\sum_{k=1}^{n}[L(x_k) - \lambda] = 0 \qquad (4)$$

Unica soluţie a ecuaţiei (4), soluţie ce depinde de valorile de selecţie este

$$\lambda = \frac{1}{n}\sum_{k=1}^{n} L(x_k)$$

ce reprezintă valoarea lui $E(X,n)$ pentru o selecţie fixată.

Teorema 2. Dacă $S(X,n)$ este o funcţie de estimaţie suficientă, atunci orice soluţie a ecuaţiei de verosimilitate este o funcţie de $S(X,n)$.

Demonstraţie. Din faptul că $S(X,n)$ este o funcţie de estimaţie suficientă, conform teoremei 1, §5, cap.IV, rezultă că

$$\ln f(x;\lambda) = A_1(\lambda)\Psi_1(x) + A_0(\lambda) + \Psi_0(x)$$

şi

$$S(X,n) = \sum_{k=1}^{n} \Psi_1(X_k)$$

sau

$$S(X,n) = \Phi\left[\sum_{k=1}^{n} \Psi_1(X_k)\right]$$

Ecuaţia de verosimilitate (2) devine

$$\sum_{k=1}^{n} \frac{\partial \ln f(x_k;\lambda)}{\partial \lambda} = 0$$

sau

$$\sum_{k=1}^{n}\left[A_1'(\lambda)\Psi_1(x_k)+A_0'(\lambda)\right]=0$$

de unde

$$(5)\qquad \sum_{k=1}^{n}\Psi_1(x_k)=-n\,\frac{A_0'(\lambda)}{A_1'(\lambda)}$$

în ipoteza că $A_1'(\lambda)\neq 0$. Orice soluţie a ecuaţiei (5) este o funcţie de $\sum_{k=1}^{n}\Psi_1(X_k)$, adică, o funcţie de $S(X,n)$.

Teorema 3. Fie $f(x;\lambda)$, $\lambda\in\Lambda\subseteq R$, densitatea de repartiţie a unei colectivităţi. Dacă sînt îndeplinite condiţiile:

a) pentru orice n există o estimaţie de verosimilitate maximă pentru λ.

b) pentru orice $\lambda\in\Lambda$ şi orice $x\in R$ există derivatele parţiale $\frac{\partial \ln f(x;\lambda)}{\partial\lambda}$, $\frac{\partial^2 \ln f(x;\lambda)}{\partial\lambda^2}$, $\frac{\partial^3 \ln f(x;\lambda)}{\partial\lambda^3}$

c) pentru orice $\lambda\in\Lambda$

$$\left|\frac{\partial f(x;\lambda)}{\partial\lambda}\right|<F_1(x),\ \left|\frac{\partial^2 f(x;\lambda)}{\partial\lambda^2}\right|<F_2(x),\ \left|\frac{\partial^3 \ln f(x;\lambda)}{\partial\lambda^3}\right|<H(x,\lambda)$$

funcţiile $F_1(x)$, $F_2(x)$ fiind integrabile pe R, iar

$$\int_R H(x,\lambda)f(x;\lambda)dx<k$$

unde k nu depinde de λ.

d) pentru orice $\lambda\in\Lambda$, $\int_R\left[\frac{\partial \ln f(x;\lambda)}{\partial\lambda}\right]^2 f(x;\lambda)dx<+\infty$,

atunci ecuaţia de verosimilitate are ca soluţie o estimaţie consistentă pentru parametrul λ. Funcţia de estimaţie corespunzătoare urmează asimptotic repartiţia normală şi este asimptotic eficientă.

Demonstraţie. A se vedea Mihoc Gh., Ciucu G., Craiu V. [10].

Metoda minimului lui hi-pătrat.

Fie $F(x;\lambda_1,\lambda_2,\ldots,\lambda_s)$, $(\lambda_1,\ldots,\lambda_s)\in\Lambda\subseteq R^s$, funcţia de repartiţie a unei colectivităţi la [illegible]

urmăreşte caracteristica X. Notăm $L = X(\Omega)$. Fie $L_i, 1 \leq i \leq k$, mulţimi boreliene ale lui R, ce realizează o partiţie a lui L. Prin urmare, au loc relaţiile

$$L = \bigcup_{i=1}^{k} L_i \ , \ L_i \cap L_j = \emptyset \ , \ i \neq j \ , \ 1 \leq i,j \leq k$$

Fie evenimentele

$$A_i = \left\{ \omega^{(n)} \in \Omega^{(n)} \ ; \ X_i(\omega^{(n)}) \in L_i \right\} \ , \ 1 \leq i \leq k \ .$$

Ţinînd seama de definiţia 3, §1, cap.III, evenimentul A_i se mai poate exprima prin

$$A_i = \left\{ \omega^{(n)} \in \Omega^{(n)} \ ; \ X(\omega_i) \in L_i \right\} \ , \ 1 \leq i \leq k \ .$$

Evident, au loc relaţiile

$$\bigcup_{i=1}^{k} A_i = \Omega^{(n)} \ , \ A_i \cap A_j = \emptyset \ , \ i \neq j \ .$$

Fie $p_i(\lambda_1,\ldots,\lambda_s) = P^{(n)}(A_i)$ şi $n_i(\omega^{(n)})$ frecvenţa absolută a evenimentului A_i în orice selecţie repetată de volum n. Din relaţiile pe care le satisfac evenimentele A_i, $1 \leq i \leq k$, rezultă că $\sum_{i=1}^{k} p_i(\lambda_1,\ldots,\lambda_s) = 1$. Mărimile $n_i(\omega^{(n)})$, $1 \leq i \leq k$, sînt variabile aleatoare. Pentru $\omega^{(n)}$ fixat, $n_i(\omega^{(n)})$ reprezintă numărul valorilor din selecţia $(x_1,\ldots,x_n)$ care aparţin mulţimii boreliene L_i. Evident, pentru $\omega^{(n)}$ fixat, are loc $\sum_{i=1}^{n} n_i(\omega^{(n)}) = n$.

<u>Definiţia 1</u>. Se spune că estimaţiile $\lambda_1^*,\ldots,\lambda_s^*$ sînt <u>estimaţii după metoda minimului lui hi-pătrat</u> respectiv pentru parametrii $\lambda_1,\ldots,\lambda_s$, dacă ele realizează minimul statisticii

$$\chi^2(\omega^{(n)};\lambda_1,..,\lambda_s) = \sum_{i=1}^{k} \frac{[n_i(\omega^{(n)}) - np_i(\lambda_1,\ldots,\lambda_s)]^2}{n\, p_i(\lambda_1,..,\lambda_s)}$$

Estimaţiile de acest fel sînt soluţii ale sistemului de ecuaţii

$$-\frac{1}{2}\frac{\partial \chi^2(\omega^{(n)};\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} = \sum_{i=1}^{k}\left[\frac{n_i-np_i}{p_i} + \frac{(n_i-np_i)^2}{2n\,p_i^2}\right]\frac{\partial p_i(\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} =$$

$$= 0, 1 \leqslant j \leqslant s$$

Chiar în cazuri simple acest sistem se rezolvă cu greutate.

Definiţia 2. Se spune că estimaţiile $\lambda_1^*,\ldots,\lambda_s^*$ sînt estimaţii după metoda modificată a minimului lui hi-pătrat respectiv pentru parametrii $\lambda_1,\ldots,\lambda_s$ dacă ele constituie o soluţie a sistemului de ecuaţii

$$(6)\qquad \sum_{i=1}^{k}\frac{n_i-np_i}{p_i}\frac{\partial p_i(\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} = 0,\ 1 \leqslant j \leqslant s$$

Din faptul că $\sum_{i=1}^{k} p_i(\lambda_1,\ldots,\lambda_s) = 1$, rezultă că

$$\sum_{i=1}^{k}\frac{\partial p_i(\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} = 0\ ,\quad 1 \leqslant j \leqslant s$$

şi de ecuaţii (6) se reduce la

$$(7)\qquad \sum_{i=1}^{k}\frac{n_i}{p_i}\frac{\partial p_i(\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} = 0\ ,\ 1 \leqslant j \leqslant s$$

Sistemul de ecuaţii (7) se mai poate scrie sub forma

$$(8)\qquad \frac{\partial}{\partial\lambda_j}\ln\prod_{i=1}^{k} p_i^{n_i} = 0\ ,\quad 1 \leqslant j \leqslant s$$

Ţinînd seama de definiţia verosimilităţii, avem

$$V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_s) = P^{(n)}(\omega^{(n)}\in\Omega^{(n)};\ Y(\omega^{(n)})=(x_1,\ldots x_n)) =$$

$$= P^{(n)}(A_1)P^{(n)}(A_2)\ldots P^{(n)}(A_k) = p_1^{n_1}p_2^{n_2}\ldots p_k^{n_k}$$

şi deci sistemul de ecuaţii (8) se scrie sub forma

$$(9)\qquad \frac{\partial \ln V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_s)}{\partial\lambda_j} = 0\ ,\ 1 \leqslant j \leqslant s$$

Teorema 5. Estimaţiile după metoda modificată a minimului lui hi-pătrat sînt estimaţii de verosimilitate maximă pentru verosimilitatea

$$V(x_1,\dots,x_n;\lambda_1,\dots,\lambda_s)=\prod_{i=1}^{k}p_i^{n_i}(\lambda_1,\dots,\lambda_s)$$

Demonstrație. Teorema rezultă din echivalența sistemelor de ecuații (6) și (9).

§ 7. ESTIMAREA PARAMETRILOR CU AJUTORUL INTERVALELOR DE ÎNCREDERE

Fie $F(x;\lambda)$ funcția de repartiție a unei colectivități, unde $\lambda\in\Lambda\subseteq R$. La estimarea punctuală a parametrului λ se căuta o funcție de estimație consistentă, ale cărei valori, pentru diferite selecții, aproximau valoarea adevărată, dar necunoscută a parametrului.

Un alt mod de estimare a parametrului λ îl constituie găsirea unui interval aleator care să acopere, cu o probabilitate foarte apropiată de 1, valoarea necunoscută a parametrului.

Presupunem că, pe o anumită cale au fost găsite statisticile $\lambda_1(X,n)$ și $\lambda_2(X,n)$, care pentru orice selecție $(x_1,\dots,x_n)\in L^n$, sînt în relația

$$\lambda_1(x_1,\dots,x_n)<\lambda_2(x_1,\dots,x_n)$$

Definiția 1. Se spune că intervalul aleator $[\lambda_1(X,n),\lambda_2(X,n)]$ constituie un interval de încredere cu nivelul de încredere δ pentru parametrul λ, dacă are loc

$$P^{(n)}(\omega^{(n)}\in\Omega^{(n)}:\lambda_1(X,n)<\lambda<\lambda_2(X,n))=\delta$$

unde δ este un număr apropiat de 1 și independent de λ.

Relația de mai înainte, ce definește intervalul de încredere, se citește astfel: „probabilitatea ca intervalul aleator $[\lambda_1(X,n),\lambda_2(X,n)]$ să acopere valoarea adevărată a parametrului λ este egală cu δ". Este greșit a citi aceeași relație astfel: „probabilitatea ca valoarea adevărată a lui λ să

se afle în intervalul $[\lambda_1(X,n), \lambda_2(X,n)]$ este egală cu δ ", întrucît parametrul λ nu este o variabilă aleatoare. Intervalul $[\lambda_1(X,n), \lambda_2(X,n)]$ este aleator întrucît extremităţile lui sînt nişte statistici. Pentru selecţia $(x_1,\ldots,x_n)$ fixată, din intervalul aleator $[\lambda_1(X,n), \lambda_2(X,)]$ se obţine intervalul determinist $[\lambda_1(x_1,\ldots,x_n), \lambda_2(x_1,\ldots,x_n)]$. Interpretarea nivelului de încredere δ este următoarea: dacă pentru N selecţii de acelaşi volum se construiesc intervalele deterministe obţinute din intervalul aleator, atunci $N \cdot \delta$ intervale acoperă valoarea adevărată a lui λ şi $N(1-\delta)$ nu acoperă. Cu cît nivelul de încredere este mai apropiat de 1 şi lungimea intervalului de încredere este mai mică, cu atît se obţine o indicaţie mai precisă cu privire la valoarea adevărată a parametrului λ.

In cele ce urmează vom prezenta două cazuri, frecvent întîlnite în aplicaţii, în care se pot construi intervale de încredere.

C a z u l I. Presupunem că există o statistică $I(X_1,\ldots,X_n;\lambda)$, notată pe scurt $I(X,n;\lambda)$, care îndeplineşte condiţiile:

a) depinde de parametrul λ pentru care se caută intervalul de încredere

b) repartiţia probabilistică a ei este complet specificată

c) este continuă şi strict monotonă în raport cu λ.

Presupunem, în plus, că repartiţia probabilistică a lui $I(X,n;\lambda)$ este continuă, dată de densitatea de repartiţie $s(x)$. Intrucît repartiţia probabilistică a lui $I(X,n;\lambda)$ este complet specificată, pentru un δ dat, cu ajutorul tabelelor acestei repartiţii probabilistice, se pot determina numerele $a(\delta)$ şi $b(\delta)$ (nu în mod unic), încît să aibă loc

(1) $P^{(n)}\{\omega^{(n)} \in \Omega^{(n)} : a(\delta) \leq I(X,n;\lambda) \leq b(\delta)\} = \delta$

ceea ce este echivalent cu

(1') $$\int_{a(\delta)}^{b(\delta)} s(x)dx = \delta$$

Evident, pentru δ dat, există o infinitate de numere $a(\delta)$ și $b(\delta)$ care să satisfacă (1').

Pentru fixarea ideilor să presupunem că $I(x_1,\dots,x_n;\lambda)$ este o funcție strict crescătoare în raport cu λ, pentru orice selecție $(x_1,\dots,x_n) \in L^n$ și orice $\lambda \in (a,b)$. Vom căuta o inegalitate echivalentă cu inegalitatea

(2) $$a(\delta) \leq I(X,n;\lambda) \leq b(\delta)$$

Pentru ușurință folosim figura 6,

Se intersectează graficul lui $y=I(x_1,\dots,x_n;\lambda)$ cu dreptele $y=a(\delta)$ și $y=b(\delta)$. Din faptul că $I(x_1,\dots,x_n;\lambda)$ este strict crescătoare în raport cu λ, rezultă că există, în mod unic, numerele $\lambda_1(x_1,\dots,x_n;\delta)$ și $\lambda_2(x_1,\dots,x_n;\delta)$ încît

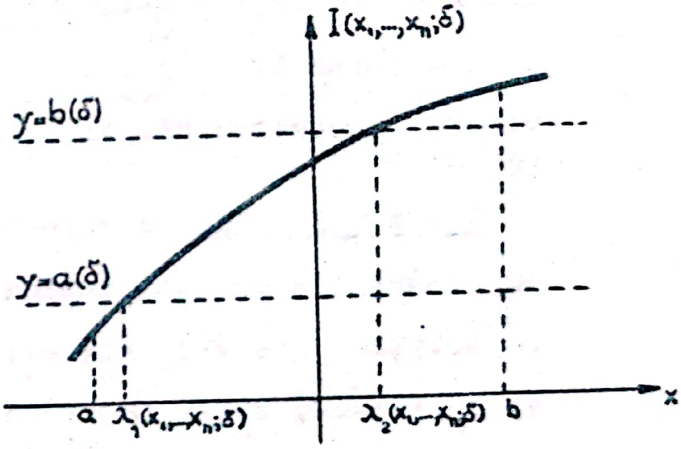


Fig.6.

(3) $$I(x_1,\dots,x_n;\lambda) \geq a(\delta)$$

să fie echivalentă cu

(3') $$\lambda \geq \lambda_1(x_1,\dots,x_n;\delta)$$

și

(4) $$I(x_1,\dots,x_n;\lambda) \leq b(\delta)$$

să fie echivalentă cu

(4') $$\lambda \leq \lambda_2(x_1,\dots,x_n;\delta)$$

Ținînd seama de relațiile (3),(3'),(4) și (4') rezultă că inegalitatea (2) este echivalentă cu inegalitatea

(5) $\lambda_1(x_1,\ldots,x_n;\delta) \leqslant \lambda \leqslant \lambda_2(x_1,\ldots,x_n;\delta)$

sau

(5') $\lambda_1(X,n;\delta) \leqslant \lambda \leqslant \lambda_2(X,n;\delta)$

Tinînd seama de echivalenţa lui (2) cu (4'), relaţia (1) devine

(6) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : \lambda_1(X,n;\delta) \leqslant \lambda \leqslant \lambda_2(X,n;\delta)) = \delta$

şi potrivit definiţiei intervalului aleator, am obţinut că $[\lambda_1(X,n;\delta), \lambda_2(X,n;\delta)]$ este un interval de încredere de nivel δ pentru parametrul λ .

După cum se vede, extremităţile intervalului de încredere depind de δ prin intermediul numerelor $a(\delta)$ şi $b(\delta)$. Intrucît pentru δ dat, numerele $a(\delta)$ şi $b(\delta)$ nu sînt determinate unic, rezultă că, pentru un acelaşi nivel de încredere δ se pot găsi o infinitate de intervale de încredere pentru parametrul λ . In cele ce urmează, vom alege pentru estimarea parametrului, intervalul de încredere de lungime minimă.

C a z u l II. Presupunem că există o statistică $H(X,n)$ cu repartiţia probabilistică continuă, dată de densitatea de repartiţie $h(x;\lambda)$. Fie k_1, k_2, numere pozitive, încît $k_1+k_2=1$ şi $a_1(\lambda,\delta)$, $b_1(\lambda,\delta)$ funcţii de λ şi δ , pentru care au loc relaţiile

(7) $$\int_a^{a_1(\lambda,\delta)} h(x;\lambda)dx = k_1(1-\delta)$$

(8) $$\int_{b_1(\lambda,\delta)}^{b} h(x;\lambda)dx = k_2(1-\delta)$$

Dacă $[a,b]$ este intervalul valorilor statisticii $H(X,n)$, atunci are loc

(9) $$\int_a^b h(x;\lambda)dx = 1$$

Din relaţiile (7),(8) şi (9), rezultă

$$\int_{a_1(\lambda,\delta)}^{b_1(\lambda,\delta)} h(x;\lambda)dx = \delta \tag{10}$$

Vom arăta că, dacă funcţia $a_1(\lambda,\delta)$ este descrescătoare în raport cu λ şi funcţia $b_1(\lambda,\delta)$ este crescătoare în raport cu λ, atunci se poate determina un interval de încredere pentru λ. Intr-adevăr, pe baza lui (10), putem scrie

$$P^{(n)}\{\omega^{(n)}\in\Omega^{(n)} : a_1(\lambda,\delta) \leqslant H(X,n) \leqslant b_1(\lambda,\delta)\} = \delta \tag{11}$$

unde δ este independent de λ. Din faptul că $a_1(\lambda,\delta)$ este descrescătoare în raport cu λ, există, aşa cum s-a demonstrat în cazul I, un număr unic $A(x_1,\ldots,x_n;\delta)$, încît inegalităţile

$$\begin{aligned} &A(x_1,\ldots,x_n;\delta) \leqslant \lambda \\ &a_1(\lambda,\delta) \leqslant H(x_1,\ldots,x_n) \end{aligned} \tag{12}$$

să fie echivalente. Analog, din faptul că $b_1(\lambda,\delta)$ este crescătoare în raport cu λ, rezultă că, există un număr unic $B(x_1,\ldots,x_n;\delta)$, încît inegalităţile

$$\begin{aligned} &\lambda \leqslant B(x_1,\ldots,x_n;\delta) \\ &H(x_1,\ldots,x_n) \leqslant b_1(\lambda,\delta) \end{aligned} \tag{13}$$

să fie echivalente. Din relaţiile (12) şi (13), rezultă că, inegalitatea

$$a_1(\lambda,\delta) \leqslant H(X,n) \leqslant b_1(\lambda,\delta) \tag{14}$$

este echivalentă cu inegalitatea

$$A(x_1,\ldots,x_n;\delta) \leqslant \lambda \leqslant B(x_1,\ldots,x_n;\delta) \tag{15}$$

Ţinînd seama de echivalenţa relaţiilor (14) şi (15), relaţia (11) devine

$$P^{(n)}\{\omega^{(n)}\in\Omega^{(n)} : A(x_1,\ldots,x_n;\delta) \leqslant \lambda \leqslant B(x_1,\ldots,x_n;\delta)\} = \delta \tag{16}$$

unde δ este independent de λ. Potrivit definiţiei intervalului de încredere, pe baza relaţiei (16), putem afirma că, $[A(X_1,\ldots,X_n;\delta),\ B(X_1,\ldots,X_n;\delta)]$ este un interval de încredere de nivel δ pentru parametrul λ.

Observaţia 1. Dacă statistica $H(X,n)$ are repartiţie probabilistică discretă, atunci în relaţiile (7) şi (8) în loc de semnul egal se va scrie semnul mai mic sau cel mult egal.

Observaţia 2. Dacă funcţia de repartiţie a colectivităţii depinde de parametrii $\lambda_1, \lambda_2, \ldots, \lambda_k$ şi se poate determina un interval de încredere pentru parametrul λ_1, oricare ar fi valorile parametrilor rămaşi $\lambda_2, \ldots, \lambda_k$, atunci parametrii $\lambda_2, \ldots, \lambda_k$ se numesc parametri inutili.

Observaţia 3. In cazul cînd trebuie estimaţi doi sau mai mulţi parametri în acelaşi timp, se poate proceda în două moduri:

a) se determină pentru fiecare parametru un interval de încredere şi deci se obţine o mulţime de intervale de încredere simultane

b) se determină o regiune de încredere care acoperă punctul parametric cu o probabilitate dată şi apropiată de 1 .

Intervalele de încredere simultane constituie un caz particular al regiunii de încredere.

Fie $F(x;\lambda_1,\ldots,\lambda_k)$ funcţia de repartiţie a unei colectivităţi, unde $(\lambda_1,\ldots,\lambda_k)\in\Lambda\subseteq R^k$. Presupunem că s-a putut determina pentru fiecare parametru λ_i, un interval de încredere
$[\lambda_1^{(i)}(X,n), \lambda_2^{(i)}(X,n)]$, cu nivelul de încredere $1-\frac{1-\delta}{k}$, $1 \leq i \leq k$.

Teorema 1. (Tukey)

Dacă $[\lambda_1^{(1)}(X,n), \lambda_2^{(1)}(X,n)], \ldots, [\lambda_1^{(k)}(X,n), \lambda_2^{(k)}(X,n)]$ sînt intervale de încredere, respectiv pentru parametrii $\lambda_1,\ldots,\lambda_k$, de nivel de încredere $1-\frac{1-\delta}{k}$, atunci probabilitatea ca aceste intervale de încredere să acopere simultan parametrii $\lambda_1,\ldots,\lambda_k$ este egală cel puţin cu δ .

Demonstrație. Fie $A_i = \{ \omega^{(n)} \in \Omega^{(n)} : \lambda_1^{(i)}(X,n) \leqslant \lambda_i \leqslant \lambda_2^{(i)}(X,n) \}$ și

$$P^{(n)}(A_i) = 1 - \frac{1-\delta}{k}, \quad 1 \leqslant i \leqslant k .$$

Evident

$$P^{(n)}(\overline{A_i}) = \frac{1-\delta}{k}, \quad 1 \leqslant i \leqslant k$$

Conform inegalității lui Boole, avem

$$P^{(n)}\left(\bigcap_{i=1}^{k} A_i\right) \geqslant 1 - \sum_{i=1}^{k} P^{(n)}(\overline{A_i}) = \delta$$

și teorema este demonstrată.

Observația 4. Uneori, pentru statistica folosită la construirea intervalului de încredere nu se cunoaște repartiția ei pentru n fixat, ci pentru $n \to \infty$. In acest caz avem de a face cu construirea intervalelor de încredere pentru selecții de volum mare.

§ 8. INTERVALUL DE ÎNCREDERE PENTRU MEDIA UNEI COLECTIVITĂȚI NORMALE IN CAZUL CÎND DISPERSIA COLECTIVITĂȚII ESTE CUNOSCUTĂ

Fie X caracteristica urmărită la indivizii unei colectivități, avînd o repartiție probabilistică $N(\lambda, \sigma^2)$ unde $\lambda = M(X)$ reprezintă parametrul ce trebuie estimat printr-un interval de încredere, iar $\sigma^2 = D^2(X)$ este cunoscut.

Conform teoremei 2, §7, cap.III, statistica

$$I(X,n;\lambda) = \frac{m(X,n) - \lambda}{\frac{\sigma}{\sqrt{n}}}$$

urmează repartiția N(0,1) . Vom arăta că această statistică îndeplinește condițiile a), b), c) din §7, cap.IV și deci cu ajutorul ei se poate determina un interval de încredere pentru parametrul λ . Evident, condițiile a) și b) sînt îndeplinite. Condiția c) este îndeplinită întrucît

$$\frac{\partial I(X,n;\lambda)}{\partial \lambda} = -\frac{1}{\frac{\sigma}{\sqrt{n}}} < 0$$

ceea ce asigură că $I(X,n;\lambda)$ este strict descrescătoare în raport cu λ.

Pentru nivelul de încredere δ fixat, se pot determina numerele $a(\delta)$ și $b(\delta)$, încît să aibă loc

$$(1) \qquad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)};\ a(\delta) < \frac{m(X,n) - \lambda}{\frac{\sigma}{\sqrt{n}}} < b(\delta)\right) = \delta$$

Relația (1) se mai poate scrie sub forma

$$(2) \quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)}; m(X,n) - \frac{\sigma}{\sqrt{n}} b(\delta) < \lambda < m(X,n) - \frac{\sigma}{\sqrt{n}} a(\delta)\right) = \delta$$

Prin urmare,

$$\left[m(X,n) - \frac{\sigma}{\sqrt{n}} b(\delta), m(X,n) - \frac{\sigma}{\sqrt{n}} a(\delta)\right]$$

este un interval de încredere pentru media colectivității. Lungimea acestui interval este

$$\frac{\sigma}{\sqrt{n}} \left[b(\delta) - a(\delta)\right]$$

Intervalul aleator de lungime minimă se obține minimizînd expresia

$$\frac{\sigma}{\sqrt{n}} \left[b(\delta) - a(\delta)\right]$$

în condiția

$$\int_{a(\delta)}^{b(\delta)} s(x)dx = \delta$$

unde $s(x)$ este densitatea de repartiție a statisticii $I(X,n;\lambda)$. Aplicînd metoda multiplicatorilor lui Lagrange, se obține $a(\delta) = -b(\delta)$. Dacă se notează $b(\delta) = z$, atunci intervalul de încredere, de lungime minimă, este

$$(3) \qquad \left[m(X,n) - \frac{\sigma}{\sqrt{n}} z, m(X,n) + \frac{\sigma}{\sqrt{n}} z\right]$$

Ținînd seama că $a(\delta) = -b(\delta)$ și $b(\delta) = z$, relația (1) devine

$$(4) \qquad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)};\ -z < \frac{m(X,n) - \lambda}{\frac{\sigma}{\sqrt{n}}} < z\right) = \delta$$

ceea ce este echivalent cu

$$2\Phi(z) = \delta \quad . \tag{4'}$$

Prin urmare, pentru δ fixat, valoarea „z" se determină cu ajutorul tabelului IV din anexă pe baza relației

$$\Phi(z) = \frac{\delta}{2} \quad . \tag{4"}$$

Tinînd seama de (3), intervalul determinist corespunzător selecției $(x_1,\dots,x_n)$ este $\left[\bar{x} - \frac{\sigma}{\sqrt{n}} z, \bar{x} + \frac{\sigma}{\sqrt{n}} z\right]$.

§ 9. INTERVALUL DE ÎNCREDERE PENTRU MEDIA UNEI COLECTIVITĂȚI NORMALE IN CAZUL CÎND DISPERSIA COLECTIVITĂȚII ESTE NECUNOSCUTĂ.

Fie X caracteristica urmărită la indivizii unei colectivități avînd repartiția probabilistică $N(\lambda, \sigma^2)$ unde $\lambda = M(X)$ reprezintă parametrul ce trebuie estimat printr-un interval de încredere, iar $\sigma^2 = D^2(X)$ este necunoscut.

Conform teoremei 9, §7, cap.III, statistica

$$I(X,n;\lambda) = \frac{m(X,n) - \lambda}{\sqrt{\frac{d_*^2(X,n)}{n}}}$$

urmează repartiția Student cu n-1 grade de libertate. Verificăm îndeplinirea condițiile a),b),c) din §7,cap.IV. Evident, condițiile a) și b) sînt îndeplinite. Condiția c) este îndeplinită, întrucît

$$\frac{\partial I(X,n;\lambda)}{\partial \lambda} = -\frac{1}{\sqrt{\frac{d_*^2(X,n)}{n}}} < 0$$

ceea ce asigură că $I(X,n;\lambda)$ este strict descrescătoare în raport cu λ .

Pentru nivelul de încredere δ fixat, se pot determina numerele $a(\delta)$ și $b(\delta)$, încît să aibă loc

$$(1)\qquad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ a(\delta)<\frac{m(X,n)-\lambda}{\sqrt{\frac{d_*^2(X,n)}{n}}}<b(\delta)\right)=\delta$$

Relaţia (1) se mai poate scrie sub forma

$$(2)\ P^{(n)}\Big(\omega^{(n)}\in\Omega^{(n)};\ m(X,n)-b(\delta)\sqrt{\frac{d_*^2(X,n)}{n}}<\lambda<m(X,n)-a(\delta)\sqrt{\frac{d_*^2(X,n)}{n}}\Big)=\delta$$

Prin urmare

$$\left[m(X,n)-b(\delta)\sqrt{\frac{d_*^2(X,n)}{n}},\ m(X,n)-a(\delta)\sqrt{\frac{d_*^2(X,n)}{n}}\right]$$

este un interval de încredere pentru media colectivităţii. Lungimea acestui interval este

$$\sqrt{\frac{d_*^2(X,n)}{n}}\left[b(\delta)-a(\delta)\right]$$

Intervalul aleator de lungime minimă se obţine minimizînd expresia

$$\sqrt{\frac{d_*^2(X,n)}{n}}\left[b(\delta)-a(\delta)\right]$$

în condiţia

$$\int_{a(\delta)}^{b(\delta)}s(x)dx=\delta$$

unde $s(x)$ este densitatea de repartiţie a statisticii $I(X,n;\lambda)$. Aplicînd metoda multiplicatorilor lui Lagrange se obţine $a(\delta)=-b(\delta)$. Dacă se notează $b(\delta)=t$, atunci intervalul de încredere de lungime minimă este

$$(3)\qquad \left[m(X,n)-t\sqrt{\frac{d_*^2(X,n)}{n}},\ m(X,n)+t\sqrt{\frac{d_*^2(X,n)}{n}}\right]$$

Ţinînd seama că $a(\delta)=-b(\delta)$ şi $b(\delta)=t$, relaţia (1) devine

$$(4)\qquad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ -t<\frac{m(X,n)-\lambda}{\sqrt{\frac{d_*^2(X,n)}{n}}}<t\right)=\delta$$

ceea ce este echivalent cu

$$(5)\quad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}:\left|\frac{m(X,n)-\lambda}{\sqrt{\frac{d_*^2(X,n)}{n}}}\right|>t\right)=1-\delta$$

Pentru δ fixat, valoarea „t" se determină cu ajutorul tabelului VII din anexă pe baza relaţiei (5). Ţinînd seama de (3), intervalul determinist corespunzător selecţiei $(x_1,\dots,x_n)$ este

$$\left[\bar{x}-t\sqrt{\frac{s_*^2}{n}}\ ,\bar{x}+t\sqrt{\frac{s_*^2}{n}}\ \right]$$

Observaţie. Ţinînd seama de observaţia 3, §7, cap.III se mai poate obţine un interval de încredere, cu ajutorul statisticii

$$I(X,n;\lambda)=\frac{m(X,n)-\lambda}{\sqrt{\frac{d^2(X,n)}{n-1}}}$$

§ 10. INTERVALUL DE ÎNCREDERE PENTRU DISPERSIA UNEI COLECTIVITĂŢI NORMALE.

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd repartiţia probabilistică $N(m,\lambda)$, $\lambda>0$, unde $m = M(X)$, iar $\sigma^2=\lambda$ reprezintă parametrul ce trebuie estimat printr-un interval de încredere.

Conform teoremei 8, §7, cap.III, statistica

$$I(X,n;\lambda)=\frac{(n-1)d_*^2(X,n)}{\lambda}$$

urmează repartiţia hi-pătrat cu n-1 grade de libertate şi parametru $\sigma=1$. Verificăm îndeplinirea condiţiilor a),b), c) din §7, cap.IV. Evident, condiţiile a) şi b) sînt îndeplinite. Condiţia c) este îndeplinită, întrucît

$$\frac{\partial I(X,n;\lambda)}{\partial\lambda}=-\frac{(n-1)d_*^2(X,n)}{\lambda^2}<0$$

ceea ce asigură că $I(X,x;\lambda)$ este strict descrescătoare în raport cu λ .

Pentru nivelul de încredere δ fixat, se pot determina numerele $a(\delta)$ şi $b(\delta)$, încît să aibă loc

$$(1)\quad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\, a(\delta)\leq\frac{(n-1)d_*^2(X,n)}{\lambda}\leq b(\delta)\right)=\delta$$

Relaţia (1) se mai poate scrie sub forma

$$(2)\quad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\, \frac{(n-1)d_*^2(X,n)}{b(\delta)}\leq\lambda\leq\frac{(n-1)d_*^2(X,n)}{a(\delta)}\right)=\delta$$

Prin urmare

$$(3)\qquad \left[\frac{(n-1)d_*^2(X,n)}{b(\delta)}\,,\ \frac{(n-1)d_*^2(X,n)}{a(\delta)}\right]$$

este un interval de încredere pentru dispersia colectivităţii, oricare ar fi valoarea medie m. Media colectivităţii este, în acest caz, un parametru inutil.

Pentru δ fixat, numerele $a(\delta)$ şi $b(\delta)$, se determină cu ajutorul tabelului V din anexă, pe baza relaţiilor

$$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}\,:\, I(X,n;\lambda)>a\right)=\frac{1+\delta}{2}$$

$$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}\,:\, I(X,n;\lambda)>b\right)=\frac{1-\delta}{2}$$

Ţinînd seama de (3), intervalul determinist corespunzător selecţiei $(x_1,\ldots,x_n)$ este

$$\left[\frac{(n-1)s_*^2}{b(\delta)}\,,\ \frac{(n-1)s_*^2}{a(\delta)}\right]$$

Observaţie. Ţinînd seama de observaţia 2, §7, cap.III se mai poate obţine un interval de încredere, cu ajutorul statisticii

$$I(X,n;\lambda)=\frac{nd^2(X,n)}{\lambda}$$

§ 11. INTERVALUL DE ÎNCREDERE PENTRU DIFERENŢA MEDIILOR A DOUĂ COLECTIVITĂŢI NORMALE ÎN CAZUL CÎND DISPERSIILE COLECTIVITĂŢILOR SÎNT CUNOSCUTE.

Fie X_1 caracteristica urmărită la indivizii colectivităţii C_1 avînd repartiţia probabilistică $N(\lambda_1,\sigma_1^2)$ şi X_2 caracteristica urmărită la indivizii colectivităţii C_2 avînd repartiţia

probabilistică $N(\lambda_2, \sigma_2^2)$ unde $\lambda_1 = M(X_1)$, $\lambda_2 = M(X_2)$, iar $\sigma_1^2 = D^2(X_1)$ şi $\sigma_2^2 = D^2(X_2)$ sînt necunoscuţi. Se caută un interval de încredere pentru $\lambda_1 - \lambda_2$.

Conform teoremei 4, §7, cap.III statistica

$$I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2) = \frac{m(X_1,n_1)-m(X_2,n_2) - (\lambda_1 - \lambda_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

urmează repartiţia $N(0,1)$. Verificăm îndeplinirea condiţiilor a),b),c) din §7, cap.IV. Evident, condiţiile a) şi b) sînt îndeplinite. Condiţia c) este îndeplinită întrucît

$$\frac{\partial I(X_1,X_2,n_1,n_2;\lambda_1 - \lambda_2)}{\partial(\lambda_1 - \lambda_2)} = - \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < 0$$

ceea ce asigură că $I(X_1,X_2,n_1,n_2;\lambda_1 - \lambda_2)$ este strict descrescătoare în raport cu $\lambda_1 - \lambda_2$. Printr-un raţionament analog celui făcut în §8, cap.IV, pentru nivelul de încredere δ fixat, se determină valoarea „z" cu ajutorul tabelului IV din anexă, pe baza relaţiei

$$\Phi(z) = \frac{\delta}{2}$$

astfel încît să aibă loc

$$(1)\quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)};\ -z \leqslant \frac{m(X_1,n_1-m(X_2,n_2)-(\lambda_1-\lambda_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leqslant z\right) = \delta$$

Relaţia (1) se mai poate scrie sub forma

$$(2)\quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)};\ m(X_1,n_1)-m(X_2,n_2)-z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leqslant \lambda_1 - \lambda_2 \leqslant\right.$$

$$\left. \leqslant m(X_1,n_1) - m(X_2,n_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \delta \quad .$$

Prin urmare

(3) $\left[m(X_1,n_1)-m(X_2,n_2)-z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, m(X_1,n_1)-m(X_2,n_2)+z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right]$

este un interval de încredere, de lungime minimă, pentru diferenţa mediilor celor două colectivităţi. Dacă $(x_1^{(1)},\dots,x_{n_1}^{(1)})$ este o selecţie din colectivitatea C_1 şi $(x_1^{(2)},\dots,x_{n_2}^{(2)})$ este o selecţie din colectivitatea C_2, atunci intervalul determinist corespunzător celor două selecţii este

$$\left[\overline{x^{(1)}}-\overline{x^{(2)}} - z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, \overline{x^{(1)}}-\overline{x^{(2)}}+z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right]$$

unde

$$\overline{x^{(1)}} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i^{(1)}, \quad \overline{x^{(2)}} = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i^{(2)}$$

§ 12. INTERVALUL DE ÎNCREDERE PENTRU DIFERENŢA MEDIILOR A DOUĂ COLECTIVITĂŢI NORMALE ÎN CAZUL CÎND DISPERSIILE COLECTIVITĂŢILOR SÎNT EGALE CU O ACEEAŞI VALOARE NECUNOSCUTĂ .

Fie X_1 caracteristica urmărită la indivizii colectivităţii C_1 avînd repartiţia probabilistică $N(\lambda_1,\sigma^2)$ şi X_2 caracteristica urmărită la indivizii colectivităţii C_2 avînd repartiţia probabilistică $N(\lambda_2,\sigma^2)$, unde $\lambda_1 = M(X_1)$, $\lambda_2 = M(X_2)$, iar $\sigma^2 = D^2(X_1) = D^2(X_2)$ este necunoscut. Se caută un interval de încredere pentru $\lambda_1 - \lambda_2$.

Conform teoremei 10, §7, cap.III, statistica

$$I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2) =$$

$$= \sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}} \; \frac{m(X_1,n_1)-m(X_2,n_2)-(\lambda_1-\lambda_2)}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}}$$

urmează repartiţia Student cu n_1+n_2-2 grade de libertate.

Verificăm îndeplinirea condiţiilor a),b) şi c) din §7, cap.IV. Evident, condiţiile a) şi b) sînt îndeplinite. Condiţia c) este îndeplinită, întrucît

$$\frac{\partial I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2)}{\partial(\lambda_1-\lambda_2)} =$$

$$= -\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\,\frac{1}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(m_2-1)d_*^2(X_2,n_2)}} < 0$$

ceea ce asigură că $I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2)$ este strict descrescătoare în raport cu $\lambda_1-\lambda_2$. Printr-un raţionament analog celui făcut în §9, cap.IV, pentru nivelul de încredere δ fixat, se determină valoarea „t", cu ajutorul tabelului VIII din anexă, pe baza relaţiei

(1) $P^{(n)}(\omega^{(n)}\in\Omega^{(n)}: |I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2)|>t) = 1-\delta$

astfel încît să aibă loc

(2) $$P^{(n)}(\omega^{(n)}\in\Omega^{(n)}:$$

$$-t \leqslant \sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\,\frac{m(X_1,n_1)-m(X_2,n_2)-(\lambda_1-\lambda_2)}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}} < t) = \delta$$

Relaţia (2) se mai poate scrie sub forma

(3) $P^{(n)}(\omega^{(n)}\in\Omega^{(n)}: m(X_1,n_1)-m(X_2,n_2)-$

$$-t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)} \leqslant$$

$$\leqslant \lambda_1-\lambda_2 \leqslant m(X_1,n_1)-m(X_2,n_2)+$$

$$+t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}) = \delta$$

Prin urmare

(4) $$\Big[m(X_1,n_1) - m(X_2,n_2) - t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2}\ ,$$

$$m(X_1,n_1) - m(X_2,n_2) + t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}\ \Big]$$

este intervalul de încredere, de lungime minimă pentru diferența mediilor celor două colectivităţi, oricare ar fi valoarea σ^2. Deci, σ^2 este un parametru inutil. Dacă $(x_1^{(1)},\dots,x_{n_1}^{(1)})$ este o selecţie din colectivitatea C_1 şi $(x_1^{(2)},\dots,x_{n_2}^{(2)})$ este o selecţie din colectivitatea C_2, atunci intervalul determinist corespunzător acestor selecţii, este

$$\Big[\bar{x}^{(1)} - \bar{x}^{(2)} - t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)s_*^{(1)2}+(n_2-1)s_*^{(2)2}}\ ,$$

$$\bar{x}^{(1)} - \bar{x}^{(2)} + t\sqrt{\frac{n_1+n_2}{n_1n_2(n_1+n_2-2)}}\sqrt{(n_1-1)s_*^{(1)2}+(n_2-1)s_*^{(2)2}}\ \Big]$$

unde

$$\bar{x}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}x_i^{(1)},\quad \bar{x}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}x_i^{(2)}$$

$$s_*^{(1)2} = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i^{(1)}-\bar{x}^{(1)})^2,\quad s_*^{(1)2} = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_i^{(2)}-\bar{x}^{(2)})^2$$

Observaţie. Ţinînd seama de observaţia 4, §7, cap.III, se mai poate obţine un interval de încredere, cu ajutorul statisticii

$$I(X_1,X_2,n_1,n_2;\lambda_1-\lambda_2) = \sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\ \frac{m(X_1,n_1)-m(X_2,n_2)-(\lambda_1-\lambda_2)}{\sqrt{n_1d^2(X_1,n_1)+n_2d^2(X_2,n_2)}}\ .$$

§ 13. INTERVALUL DE ÎNCREDERE PENTRU RAPORTUL DISPERSIILOR A DOUĂ COLECTIVITĂȚI NORMALE.

Fie X_1 caracteristica urmărită la indivizii colectivității C_1 având repartiția probabilistică $N(m_1, \lambda_1)$ și X_2 caracteristica urmărită la indivizii colectivității C_2 având repartiția probabilistică $N(m_2, \lambda_2)$, unde $m_1 = M(X_1)$, $\lambda_1 = D^2(X_1)$, $m_2 = M(X_2)$, $\lambda_2 = D^2(X_2)$. Se caută un interval de încredere pentru $\frac{\lambda_1}{\lambda_2}$.

Conform teoremei 11, §7, cap.III, statistica

$$I(X_1, X_2, n_1, n_2; \frac{\lambda_1}{\lambda_2}) = \frac{\lambda_2}{\lambda_1} \frac{d_*^2(X_1, n_1)}{d_*^2(X_2, n_2)}$$

urmează repartiția Snedecor cu n_1-1 și n_2-1 grade de libertate. Verificăm îndeplinirea condițiilor a), b), c) din §7, cap.IV. Evident, condițiile a) și b) sînt îndeplinite. Condiția c) este îndeplinită, întrucît

$$\frac{\partial I(X_1, X_2, n_1, n_2; \frac{\lambda_1}{\lambda_2})}{\partial \frac{\lambda_1}{\lambda_2}} = -(\frac{\lambda_2}{\lambda_1})^2 \frac{d_*^2(X_1, n_1)}{d_*^2(X_2, n_2)} < 0$$

ceea ce asigură că $I(X_1, X_2, n_1, n_2; \frac{\lambda_1}{\lambda_2})$ este strict descrescătoare în raport cu $\frac{\lambda_1}{\lambda_2}$.

Pentru nivelul de încredere δ fixat, se pot determina numerele $a(\delta)$ și $b(\delta)$, încît să aibă loc

$$(1)\quad P^{(n)}\{\omega^{(n)} \in \Omega^{(n)};\ a(\delta) < \frac{\lambda_2 d_*^2(X_1, n_1)}{\lambda_1 d_*^2(X_2, n_2)} < b(\delta)\} = \delta$$

Relația (1) se mai poate scrie sub forma

(2) $P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}\,;\ \frac{d_*^2(X_1,n_1)}{b(\delta)d_*^2(X_2,n_2)}\leqslant\frac{\lambda_1}{\lambda_2}\leqslant\frac{d_*^2(X_1,n_1}{a(\delta)d_*^2(X_2,n_2)}\right)=\delta$

Prin urmare

(3) $$\left[\frac{d_*^2(X_1,n_1)}{b(\delta)d_*^2(X_2,n_2)}\,,\frac{d_*^2(X_1,n_1)}{a(\delta)d_*^2(X_2,n_2)}\right]$$

este un interval de încredere pentru raportul dispersiilor celor două colectivităţi normale, oricare ar fi valorile medii m_1,m_2. Rezultă că, m_1 şi m_2 sînt parametri inutili.

Pentru δ fixat, numerele $a(\delta)$ şi $b(\delta)$, se determină cu ajutorul tabelelor VIII sau IX din anexă, pe baza relaţiilor

$$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ I(X_1,X_2,n_1,n_2;\frac{\lambda_1}{\lambda_2})>a\right)=\frac{1+\delta}{2}$$

$$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ I(X_1,X_2,n_1,n_2;\frac{\lambda_1}{\lambda_2})>b\right)=\frac{1-\delta}{2}$$

Dacă $(x_1^{(1)},\ldots,x_{n_1}^{(1)})$ este o selecţie din colectivitatea C_1 şi $(x_1^{(2)},\ldots,x_{n_2}^{(2)})$ este o selecţie din colectivitatea C_2, atunci ţinînd seama de (3), intervalul determinist corespunzător acestor selecţii este

$$\left[\frac{s_*^{(1)2}}{b(\delta)s_*^{(2)2}},\ \frac{s_*^{(1)2}}{a(\delta)s_*^{(2)2}}\right].$$

Observaţie. Ţinînd seama de observaţia 5, §7, cap.III, se mai poate obţine un interval de încredere, cu ajutorul statisticii

$$I(X_1,X_2,n_1,n_2;\frac{\lambda_1}{\lambda_2})=\frac{n_1(n_2-1)d_*^2(X_1,n_1)}{n_2(n_1-1)d_*^2(X_2,n_2)}\,.$$

§ 14. INTERVALE DE ÎNCREDERE PENTRU SELECȚII MARI.

Intervalele de încredere, de care ne-am ocupat în paragrafele precedente, au fost determinate pentru un volum de selecție fixat. In acest paragraf, vom arăta că, în condiții destul de largi, se poate găsi un interval de încredere pentru selecții mari, folosind verosimilitatea.

Fie $f(x;\lambda)$ densitatea de repartiție a caracteristicii X și verosimilitatea $V(x_1,\dots,x_n;\lambda) = \prod_{k=1}^{n} f(x_k;\lambda)$. Evident,

$$\frac{\partial \ln V(x_1,\dots,x_n;\lambda)}{\partial\lambda} = \sum_{k=1}^{n} \frac{\partial \ln f(x_k;\lambda)}{\partial\lambda}$$

Dacă se notează

$$Y_i = \frac{\partial \ln f(X_i;\lambda)}{\partial\lambda}, \quad 1 \leqslant i \leqslant n,$$

atunci, avem

$$\frac{\partial \ln V(X_1,\dots,X_n;\lambda)}{\partial\lambda} = \sum_{k=1}^{n} Y_k \tag{1}$$

Din faptul că, $f(x;\lambda)$ este densitate de repartiție, rezultă

$$\int_{-\infty}^{+\infty} f(x;\lambda)dx = 1 . \tag{2}$$

Derivînd relația (2) în raport cu λ, în presupunerea că acest lucru este posibil, se obține

$$\int_{-\infty}^{+\infty} \frac{\partial f(x;\lambda)}{\partial\lambda}dx = 0 \tag{3}$$

Dacă $Y = \frac{\partial \ln f(X;\lambda)}{\partial\lambda}$ este o variabilă aleatoare, atunci

$$M(Y) = \int_{-\infty}^{+\infty} \frac{\partial \ln f(x;\lambda)}{\partial\lambda} f(x;\lambda)dx = \int_{-\infty}^{+\infty} \frac{\partial f(x;\lambda)}{\partial\lambda}dx . \tag{4}$$

Tinînd seama de (3), relația (4) devine

$$M(Y) = 0 . \tag{5}$$

Presupunem că sînt îndeplinite condițiile din teorema lui

Leapunov, §1, cap.I. Fie $D^2(Y) = D^2$. Ținînd seama că variabilele aleatoare Y_i, $1 \leqslant i \leqslant n$, sînt independente și au aceeași repartiție ca și variabila aleatoare Y, rezultă

$$M(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} M(Y_i) = 0$$

$$D^2(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} D^2(Y_i) = nD^2$$

Fie statistica

$$I(X,n;\lambda) = \frac{Y_1+\ldots+Y_n}{\sqrt{n}\,D} .$$

Ținînd seama de (1), statistica $I(X,n;\lambda)$ devine

$$I(X,n;\lambda) = \frac{1}{\sqrt{n}\,D} \frac{\partial \ln V(X_1,\ldots,X_n;\lambda)}{\partial \lambda}$$

Conform teoremei lui Leapunov, §1, cap.I, statistica $I(X,n;\lambda)$ urmează asimptotic repartiția $N(0,1)$. Dacă, în plus, $I(X,n;\lambda)$ este strict monotonă în raport cu λ, atunci se poate obține pentru parametrul λ un interval de încredere, după procedeul indicat în cazul I, §7, cap.IV.

Observație. Cele spuse mai înainte rămîn valabile și în cazul cînd caracteristica X are o repartiție probabilistică discretă.

P R O B L E M E

1. Un eveniment A se realizează în fiecare probă cu probabilitatea p. Fie $X(\omega)$ numărul de realizări ale lui A în n probe independente. Să se arate că statistica $\frac{X(\omega)}{n}$ este o funcție de estimație absolut corectă pentru p, iar statisticile $\frac{X(\omega)}{n+1}$ și $\frac{X(\omega)+1}{n+1}$ sînt funcții de estimație corectă pentru p.

2. Caracteristica X a unei colectivități urmează o repartiție normală de medie m cunoscută și dispersie λ necunoscută.

Să se arate că statistica

$$S(X,n) = \frac{1}{n}\sqrt{0,5\pi}\sum_{k=1}^{n}|X_k - m|$$

este o funcţie de estimaţie corectă pentru λ , unde X_k , $1 \leqslant k \leqslant n$, sînt variabilele aleatoare de selecţie.

3. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$P(X = k) = C_s^k \lambda^k (1-\lambda)^{s-k} \quad , \quad k = 0,1,\ldots,s$$

Se cere:

a) să se găsească o funcţie de estimaţie pentru λ .

b) să se determine natura funcţiei de estimaţie găsite la punctul precedent

c) să se afle estimaţia lui λ folosind selecţia

$$x_1 = 1 \ , \ x_2 = 2 \ , \ x_3 = 3 \ .$$

4. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad , \quad k = 0,1,2,\ldots$$

Se cere:

a) să se găsească o funcţie de estimaţie pentru λ

b) să se determine natura funcţiei de estimaţie găsite la punctul precedent

c) să se afle estimaţia lui λ folosind selecţia

$$x_1 = 2 \ , \ x_2 = 3 \ , \ x_3 = 5 \ , \ x_4 = 6 \ .$$

5. Repartiţia unei colectivităţi este dată de

$$f(x) = \begin{cases} \dfrac{x^{p-1}\, e^{-\frac{x}{\lambda}}}{\Gamma(p)\,\lambda^p} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

unde p este cunoscut.

Se cere:

a) să se găsească o funcţie de estimaţie pentru λ.

b) să se determine natura funcţiei de estimaţie găsite la punctul precedent

c) să se afle estimaţia lui λ folosind selecţia

$x_1 = 1,3 \quad x_2 = 1,7 \quad x_3 = 1,4 \quad x_4 = 0,8$

6) Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad , \quad k = 0,1,2,\ldots$$

Să se arate că media de selecţie este o funcţie de estimaţie suficientă pentru λ .

7. Repartiţia unei colectivităţi este dată de

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}\,(x-c)}\, e^{-\frac{[\ln(x-c)-\lambda]^2}{2\sigma^2}} & x > c \\ 0 & x \leqslant c \end{cases}$$

unde σ este cunoscut.

Se cere:

a) să se găsească o funcţie de estimaţie pentru λ

b) să se determine natura funcţiei de estimaţie găsite la punctul precedent

8. Repartiţia unei colectivităţi este dată de

$$f(x) = \begin{cases} \frac{1}{\lambda}\, e^{-\frac{x}{\lambda}} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

Se cere:

a) să se găsească o funcţie de estimaţie pentru λ

b) să se determine natura funcţiei de estimaţie găsite la punctul precedent

c) să se afle estimaţia lui λ folosind selecţia

$x_1 = 2,3 \quad x_2 = 3,7 \quad x_3 = 3$.

9. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$P(X = k) = C_n^k p^k (1-p)^{n-k} , \quad k = 0,1,2,\dots,n .$$

Să se estimeze p printr-un interval de încredere de nivel de încredere $\delta = 0,96$ folosind selecţia

$$x_1 = 1 \qquad x_2 = 2 \qquad x_3 = 3 .$$

10. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Să se estimeze λ printr-un interval de încredere de nivel de încredere $\delta = 0,95$ folosind selecţia

$$x_1 = 2,3 \qquad x_2 = 1,2 \qquad x_3 = 0,9 \qquad x_4 = 0,2$$

11. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$X : \begin{bmatrix} & 1 \\ q & p \end{bmatrix} , \quad p+q = 1 .$$

Să se determine pentru parametrul p, un interval de încredere, în cazul $n \to \infty$.

12. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} , \quad k = 0,1,2,\dots$$

Să se determine pentru parametrul λ , un interval de încredere, în cazul $n \to \infty$.

13. Concentraţia în pentaoxid de fosfor, dintr-un anumit îngrăşământ, este o caracteristică ce urmează repartiţia normală de medie m necunoscută şi dispersie $\sigma^2 = 3,61$. Să se estimeze concentraţia medie în pentaoxid de fosfor, printr-un interval de încredere cu nivelul de încredere $\delta = 0,95$, folosind selecţia

$x_1 = 19,4$ $x_2 = 22,2$ $x_3 = 21,9$ $x_4 = 23,2$ $x_5 = 19,8$ $x_6 = 21,3$
$x_7 = 17,8$ $x_8 = 23,2$ $x_9 = 18,7$ $x_{10} = 19,3$ $x_{11} = 18,9$ $x_{12} = 19,2$

14. Rezistența la rupere a unui fir de bumbac este o caracteristică ce urmează o repartiție normală de parametri m și σ^2 necunoscuți. Să se estimeze rezistența medie a firului printr-un interval de încredere cu nivelul de încredere $\delta = 0,97$ folosind selecția

$$x_1 = 63 \quad x_2 = 71 \quad x_3 = 52 \quad x_4 = 49 .$$

15. Conținutul de oxizi metalici în soda caustică este o caracteristică ce urmează o repartiție normală de parametri m σ^2 necunoscuți. Să se estimeze dispersia conținutului de oxizi metalici printr-un interval de încredere cu nivelul de încredere $\delta = 0,95$ folosind selecția

$x_1 = 0,166$ $x_2 = 0,144$ $x_3 = 0,128$ $x_4 = 0,044$ $x_5 = 0,134$ $x_6 = 0,140$.

16. Înălțimea persoanelor unui grup este o caracteristică ce urmează o repartiție normală de medie m și dispersie σ^2. Să se estimeze parametrii m și σ^2 printr-o regiune de încredere cu nivelul de încredere $\delta = 0,97$ folosind datele din tabelul 3

Tabelul 3

Înălțimea în cm.	până la 100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	peste 170
Frecvența absolută	3	8	13	21	32	24	12	6	1

C a p i t o l u l V

VERIFICAREA IPOTEZELOR STATISTICE

$x_1 = 19,4$ $x_2 = 22,2$ $x_3 = 21,9$ $x_4 = 23,2$ $x_5 = 19,8$ $x_6 = 21,3$
$x_7 = 17,8$ $x_8 = 23,2$ $x_9 = 18,7$ $x_{10} = 19,3$ $x_{11} = 18,9$ $x_{12} = 19,2$

14. Rezistența la rupere a unui fir de bumbac este o caracteristică ce urmează o repartiție normală de parametri m și σ^2 necunoscuți. Să se estimeze rezistența medie a firului printr-un interval de încredere cu nivelul de încredere $\delta = 0,97$ folosind selecția

$$x_1 = 63 \quad x_2 = 71 \quad x_3 = 52 \quad x_4 = 49 .$$

15. Conținutul de oxizi metalici în soda caustică este o caracteristică ce urmează o repartiție normală de parametri m σ^2 necunoscuți. Să se estimeze dispersia conținutului de oxizi metalici printr-un interval de încredere cu nivelul de încredere $\delta = 0,95$ folosind selecția

$x_1 = 0,166$ $x_2 = 0,144$ $x_3 = 0,128$ $x_4 = 0,044$ $x_5 = 0,134$ $x_6 = 0,140$.

16. Înălțimea persoanelor unui grup este o caracteristică ce urmează o repartiție normală de medie m și dispersie σ^2. Să se estimeze parametrii m și σ^2 printr-o regiune de încredere cu nivelul de încredere $\delta = 0,97$ folosind datele din tabelul 3

Tabelul 3

Înălțimea în cm.	până la 100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	peste 170
Frecvența absolută	3	8	13	21	32	24	12	6	1

Scanned with OKEN Scanner

C a p i t o l u l V

VERIFICAREA IPOTEZELOR STATISTICE

Scanned with OKEN Scanner

§ 1. GENERALITĂŢI ASUPRA IPOTEZELOR STATISTICE

In general, funcţia de repartiţie a unei colectivităţi poate fi numai specificată sau complet necunoscută. Pe baza repartiţiilor statistice asociate diferitelor selecţii efectuate din colectivitate, se ajunge la formulareaanumitor presupuneri, fie asupra valorilor parametrilor necunoscuţi ce intră în funcţia de repartiţie a colectivităţii, fie asupra formei funcţionale a acesteia.

Definiţia 1. Se numeşte ipoteză statistică orice presupunere cu privire la funcţia de repartiţie specificată sau necunoscută a unei colectivităţi.

Se disting două feluri de ipoteze statistice: parametrice şi neparametrice.

Definiţia 2. Se numeşte ipoteză statistică parametrică o ipoteză statistică cu privire la parametri ce intră în funcţia de repartiţie a colectivităţii.

Definiţia 3. Se numeşte ipoteză statistică neparametrică o ipoteză statistică cu privire la forma funcţională a funcţiei de repartiţie a colectivităţii.

Fie $F(x;\lambda_1,\dots,\lambda_k)$, unde $\lambda=(\lambda_1,\dots,\lambda_k)\in\Lambda\subseteq R^k$, funcţia de repartiţie a unei colectivităţi. Punctul parametric $\lambda=(\lambda_1,\dots,\lambda_k)$ are în spaţiulparametrilor Λ o poziţie bine determinată, dar această poziţie este necunoscută. Ipotezele parametrice pot fi simple sau compuse.

Definiția 4. Se numește ipoteză parametrică simplă o ipoteză parametrică ce specifică valorile tuturor parametrilor necunoscuți.

Exemplu :

$$H : \lambda_1 = \lambda_1^0, \ldots, \lambda_k = \lambda_k^0$$

Ipoteza H selecționează din spațiul parametrilor Λ punctul $(\lambda_1^0, \ldots, \lambda_k^0)$.

Definiția 5. Se numește ipoteză parametrică compusă o ipoteză parametrică ce specifică valorile a numai ℓ parametri ($\ell < k$) din cei k parametri necunoscuți.

Exemplu:

$$H : \lambda_1 = \lambda_1^0, \ldots, \lambda_\ell = \lambda_\ell^0$$

Ipoteza H selecționează din spațiul parametrilor Λ un hiperplan $k - \ell$ dimensional.

Fie $F(x; \lambda)$ funcția de repartiție a unei colectivități, unde $\lambda \in \Lambda \subseteq R$, și ipoteza parametrică simplă

$$H : \lambda = \lambda_0 .$$

Dacă parametrul λ poate lua și alte valori diferite de λ_0, atunci ipoteza .

$$H' : \lambda \neq \lambda_0$$

se numește ipoteză alternativă a ipotezei H. O ipoteză statistică poate avea una sau mai multe alternative.

§ 2. TEST, FUNCȚIE DE PUTERE, FUNCȚIE CARACTERISTICĂ OPERATOARE

Intrucît o ipoteză statistică este o presupunere făcută pe baza repartițiilor statistice ale colectivității, această presupunere poate fi în realitate adevărată sau falsă. De aici rezultă necesitatea testării ipotezei formulate.

A testa o ipoteză statistică înseamnă a lua una din deciziile

i) ipoteza se respinge
ii) ipoteza se acceptă

Definiția 1. Se numește test statistic orice procedeu de verificare a unei ipoteze statistice.

Orice test statistic se bazează pe un criteriu.

Definiţia 2. Se numeşte criteriu pentru testarea unei ipoteze statistice, o statistică ce îndeplineşte condiţiile :

a) depinde de ipoteza făcută

b) dacă ipoteza făcută se acceptă, atunci repartiţia probabilistică a ei este complet specificată.

Vom nota criteriul de testare prin $C(X,n)$. Testarea unei ipoteze statistice se face la un anumit nivel de semnificaţie.

Definiţia 3. Se numeşte nivel de semnificaţie probabilitatea respingerii ipotezei făcute în cazul cînd ea în realitate este adevărată.

Vom nota nivelul de smnificaţie cu α . Pentru ca ipoteza adevărată să fie respinsă în cît mai puţine cazuri, nivelul de semnificaţie se alege apropiat de zero. De obicei, $\alpha = 0,05$; $\alpha = 0,01$.

Fie B o mulţime boreliană din spaţiul parametrilor Λ , ipoteza

$$H : (\lambda_1,\dots,\lambda_k) \in B$$

ipoteza alternativă

$$H' : (\lambda_1,\dots,\lambda_k) \notin B$$

şi $T(X,n)$ statistica obţinută din $C(X,n)$ după acceptarea lui H. Intrucît repartiţia probabilistică a lui $T(X,n)$ este complet specificată, pentru nivelul de semnificaţie α fixat, se poate determina (nu în mod unic) o mulţime $I \subset R$, încît să aibă loc

(1) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : T(X,n) \in I \mid H \text{ se acceptă}) = \alpha$

dacă $T(X,n)$ urmează o repartiţie probabilistică continuă. Mulţimii I îi va corespunde în spaţiul valorilor de selecţie L^n o mulţime U , astfel încît

(2) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \in U \mid H \text{ se acceptă}) = \alpha$

unde $Y(\omega^{(n)})$ este vectorul de selecţie al caracteristicii X.

Observaţie. Dacă $T(X,n)$ urmează o repartiţie probabilistică discretă, atunci în relaţiile (1) şi (2) se va scrie în locul semnului egal, semnul mai mic sau cel mult egal.

Mulţimea U din L^n definită prin (2) se numeşte regiune critică. Pentru un α fixat se pot găsi mai multe mulţimi U ce satisfac (2). Fiecare regiune critică furnizează un test statistic. Prin urmare, pentru un nivel de semnificaţie α fixat, pot exista mai multe teste pentru verificarea aceleiaşi ipoteze. Presupunem că, atunci cînd T(X,n) variază în I, vectorul de selecţie $Y(\omega^{(n)})$ variază în U şi reciproc. Dacă nivelul de semnificaţie este α şi se efectuează N selecţii de acelaşi volum, atunci, în medie, $N\alpha$ selecţii cad în U. Considerînd că într-o singură probă, căderea punctului $(x_1,\ldots,x_n)$ în U, este un eveniment rar, putem conveni a respinge ipoteza H, cînd acest lucru se întîmplă.

Fie T_o valoarea statisticii T(X,n) pentru selecţia pe baza căreia se testează ipoteza. Facem următoarea convenţie:

i) dacă $(x_1,\ldots,x_n)\in U$, ceea ce este echivalent cu $T_o\in I$, ipoteza H se respinge

ii) dacă $(x_1,\ldots,x_n)\notin U$, ceea ce este echivalent cu $T_o\notin I$, ipoteza H se acceptă, pînă la momentul unei testări mai puternice.

Această convenţie constituie testul de verificare. Intrucît respingerea sau acceptarea unei ipoteze statistice este o convenţie, la testarea unei ipoteze se pot comite erori. Se pot întîmpla următoarele situaţii :

i) deşi $(x_1,\ldots,x_n)\in U$, ceea ce este echivalent cu $T_o\in I$, ipoteza H să fie adevărată. Aplicînd testul de mai înainte, ipoteza H se respinge. Eroarea care se comite respingînd o ipoteză adevărată se numeşte eroare de genul I. Probabilitatea comiterii unei erori de genul I se numeşte risc de genul I. Tinînd seama de definiţia 3, riscul de genul I este tocmai nivelul de semnificaţie.

ii) deşi $(x_1,\ldots,x_n)\notin U$, ceea ce este echivalent cu $T_o\notin I$,

ipoteza H să fie falsă. Aplicînd testul de mai înainte ipoteza H se acceptă. Eroarea care se comite acceptînd o ipoteză falsă se numeşte <u>eroare de genul II</u>. Probabilitatea comiterii unei erori de genul II se numeşte <u>risc de genul II</u>. Riscul de genul II îl vom nota cu β .

Ţinînd seama de definiţiile riscurilor, se poate scrie

$$\alpha = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \in U \mid H \text{ se acceptă})$$

(3)

$$\beta = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \notin U \mid H' \text{ se acceptă})$$

Nu există o regulă conform căreia să fie de preferat unul dintre riscuri. Dacă se fixează unul dintre riscuri, celălalt este funcţie de cel fixat.

Testarea unei ipoteze statistice se efectuează după următoarea schemă:

1. Alegerea criteriului C(X,n), care după acceptarea ipotezei de testat, devine T(X,n).

2. Alegerea nivelului de semnificaţie α foarte apropiat de zero.

3. Determinarea mulţimii I, astfel încît să aibă loc

$$P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : T(X,n) \in I \mid H \text{ se acceptă}) = \alpha$$

4. Calcularea valorii T_o a lui T(X,n) pentru selecţia pe baza căreia se face testarea.

5. Luarea deciziei (testul)

i) dacă $T_o \in I$, ipoteza H se respinge

ii) dacă $T_o \notin I$, ipoteza H se acceptă

Fie $F(x;\lambda)$, $\lambda \in \Lambda \subseteq R$, funcţia de repartiţie a unei colectivităţi. Presupunem că trebuie testată ipoteza

$$H_o : \lambda = \lambda_o$$

faţă de ipoteza alternativă

$$H_1 : \lambda = \lambda_1$$

<u>Definiţia 4</u>. Se numeşte <u>funcţie de putere</u> a unui test pentru verificarea ipotezei H_o, probabilitatea respingerii lui H_o.

Dacă testul se bazează pe regiunea critică U şi ipoteza se referă la parametrul λ , atunci funcţia de putere o vom nota prin $\bar{\pi}(U,\lambda)$. Ţinînd seama de definiţia 4 şi relaţiile (3), avem

$$\bar{\pi}(U,\lambda_0)=P^{(n)}(\omega^{(n)}\in\Omega^{(n)};\ Y(\omega^{(n)})\in U|H_0 \text{ se acceptă})=\alpha$$

(4) $$\bar{\pi}(U,\lambda_1)=P^{(n)}(\omega^{(n)}\in\Omega^{(n)};\ Y(\omega^{(n)})\in U|H_1 \text{ se acceptă})=$$

$$=1-P^{(n)}(\omega^{(n)}\in\Omega^{(n)};Y(\omega^{(n)})\notin U|H_1 \text{ se acceptă})=1-\beta.$$

Definiţia 5. Se numeşte funcţie caracteristică operatoare a unui test pentru verificarea ipotezei H_0, probabilitatea acceptării lui H_0.

Vom nota funcţia caracteristică operatoare a unui test bazat pe regiunea critică U prin $L(U,\lambda)$. Ţinînd seama de definiţia 5 şi relaţiile (4), rezultă

(5) $$L(U,\lambda_0)=1-\bar{\pi}(U,\lambda_0)=1-\alpha\ ,$$
$$L(U,\lambda_1)=1-\bar{\pi}(U,\lambda_1)=\beta\ .$$

Definiţia 6. Se numeşte test ideal pentru verificarea ipotezei H_0 un test pentru care

$$L(U,\lambda)=\begin{cases}1 & \text{dacă } \lambda=\lambda_0\\ 0 & \text{dacă } \lambda\neq\lambda_0\end{cases}$$

Cu alte cuvinte, pentru un test ideal, probabilitatea acceptării ipotezei H_0 cînd ea este adevărată este egală cu unu, iar probabilitatea respingerii ipotezei H_0 cînd ea este falsă este egală cu zero. Un test ideal nu se poate obţine.

§ 3. CEL MAI PUTERNIC TEST. LEMA NEYMAN-PEARSON.

Presupunem că trebuie testată ipoteza $H_0:\lambda=\lambda_0$ faţă de ipoteza alternativă $H_1:\lambda\neq\lambda_0$, la nivelul de semnificaţie α.

Definiţia 1. Se spune că testul bazat pe regiunea critică U^* este cel mai puternic în raport cu toate testele bazate pe regiunile critice U, la nivelul de semnificaţie α, dacă sînt îndeplinite condiţiile :

a) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U^* \mid H_0 \text{ se acceptă}) = \alpha$

b) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U^* \mid H_1 \text{ se acceptă}) \geqslant$

$\geqslant P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă})$

pentru orice regiune critică U.

Calculînd riscul de genul II pentru regiunea critică U^* se obţine

(1) $\beta_{U^*} = P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \notin U^* \mid H_1 \text{ se acceptă}) =$

$= 1 - P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U^* \mid H_1 \text{ se acceptă})$

Ţinînd seama de condiţia b) din definiţia 1, rezultă

(2) $1 - P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U^* \mid H_1 \text{ se acceptă}) \leqslant$

$\leqslant 1 - P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă}) =$

$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)};\ Y(\omega^{(n)}) \notin U \mid H_1 \text{ se acceptă})$

Din (1) şi (2), rezultă

(3) $\beta_{U^*} \leqslant \beta_U$

oricare ar fi regiunea critică U.

Pe baza relaţiei (3) se poate afirma că, dintre toate testele de nivel de semnificaţie α fixat, cel mai puternic este acela pentru care riscul de genul al doilea este cel mai mic. Nu totdeauna există un test cel mai puternic.

Presupunem că parametrul λ din funcţia de repartiţie $F(x;\lambda)$ poate lua numai valorile λ_0 şi λ_1. Vom arăta că, în anumite condiţii, există un test cel mai puternic pentru testarea ipotezei $H_0 : \lambda = \lambda_0$ faţă de ipoteza alternativă $H_1 : \lambda = \lambda_1$.

LEMA NEYMAN - PEARSON

Fie $P(x_1,\ldots,x_n;\lambda)$ legea de probabilitate a vectorului de selecţie $Y(\omega^{(n)})$, unde

$$P(x_1,\ldots,x_n;\lambda) = \begin{cases} P(x_1,\ldots,x_n;\lambda_0) \text{ dacă } H_0 \text{ se acceptă} \\ P(x_1,\ldots,x_n;\lambda_1) \text{ dacă } H_1 \text{ se acceptă} \end{cases}$$

Fie $\mathcal{U}$ mulţimea tuturor regiunilor critice de nivel de semnificaţie α pentru verificarea ipotezei H_0 faţă de ipoteza

alternativă H_1 . Dacă există o regiune critică $U^* \in \mathcal{U}$, în-cît pentru $k > 0$, să aibă loc relaţiile

$$P(x_1,\dots,x_n;\lambda_1) \geqslant kP(x_1,\dots,x_n;\lambda_o) \quad \text{dacă } (x_1,\dots,x_n) \in U^*$$

$$P(x_1,\dots,x_n;\lambda_1) \leqslant kP(x_1,\dots,x_n;\lambda_o) \quad \text{dacă } (x_1,\dots,x_n) \in U^{*c}$$

atunci U^* este regiunea critică pe care se bazează cel mai puternic test de nivel de semnificaţie α pentru testarea ipotezei H_o faţă de ipoteza alternativă H_1 .

Demonstraţie. Vom demonstra lema în cazul cînd $Y(\omega^{(n)})$ urmează o repartiţie probabilistică discretă. Mulţimea $\mathcal{U}$ a regiunilor critice, de nivel de semnificaţie α , este

$$\mathcal{U} = \{U \subset L^n : P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; Y(\omega^{(n)}) \in U \mid H_o \text{ se acceptă}) = \alpha\}$$

Intrucît $Y(\omega^{(n)})$ urmează o repartiţie probabilistică discretă, rezultă

$$(1)\quad P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; Y(\omega^{(n)}) \in U \mid H_o \text{ se acceptă}) = \sum_{(x_1,\dots,x_n)\in U} P(x_1,\dots,x_n;\lambda_o) .$$

Din faptul că $U \in \mathcal{U}$ şi din (1), urmează

$$(2)\quad \sum_{(x_1,\dots,x_n)\in U} P(x_1,\dots,x_n;\lambda_o) = \alpha$$

Analog, din faptul că $U^* \in \mathcal{U}$ şi din (1), urmează

$$(3)\quad \sum_{(x_1,\dots,x_n)\in U^*} P(x_1,\dots,x_n;\lambda_o) = \alpha$$

A demonstra că testul bazat pe regiunea critică U^* este cel mai puternic, potrivit definiţiei 1, revine la a arăta că are loc

$$(4)\quad P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; Y(\omega^{(n)}) \in U^* \mid H_1 \text{ se acceptă}) \geqslant P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă})$$

- oricare ar fi $U \in \mathcal{U}$. Ţinînd seama de (1), relaţia (4) devine

$$(5)\quad \sum_{(x_1,\dots,x_n)\in U^*} P(x_1,\dots,x_n;\lambda_1) \geqslant \sum_{(x_1,\dots,x_n)\in U} P(x_1,\dots,x_n;\lambda_1)$$

oricare ar fi $U \in \mathcal{U}$.

Dacă se notează

(6) $$\delta = \sum_{(x_1,..,x_n)\in U^*} P(x_1,...,x_n;\lambda_1) - \sum_{(x_1,..,x_n)\in U} P(x_1,...,x_n;\lambda_1)$$

atunci avem de demonstrat că $\delta \geqslant 0$, oricare ar fi $U \in \mathcal{U}$.

Avînd în vedere relaţiile

$$U \cup U^c = L^n \quad , \quad U \cap U^c = \emptyset$$

$$U^* \cup U^{*c} = L^n \quad , \quad U^* \cap U^{*c} = \emptyset$$

sumele din membrul drept al relaţiei (6) pot fi exprimate prin

(7) $$\sum_{(x_1,..,x_n)\in U^*} P(x_1,...,x_n;\lambda_1) = \sum_{(x_1,..,x_n)\in U^* \cap L^n} P(x_1,...,x_n;\lambda_1) =$$

$$= \sum_{(x_1,..,x_n)\in U^* \cap (U \cup U^c)} P(x_1,...,x_n;\lambda_1) = \sum_{(x_1,..,x_n)\in U^* \cap U} P(x_1,...,x_n;\lambda_1) +$$

$$+ \sum_{(x_1,..,x_n)\in U^* \cap U^c} P(x_1,...,x_n;\lambda_1)$$

şi

(8) $$\sum_{(x_1,..,x_n)\in U} P(x_1,...,x_n;\lambda_1) = \sum_{(x_1,..,x_n)\in U \cap L^n} P(x_1,...,x_n;\lambda_1) =$$

$$= \sum_{(x_1,..,x_n)\in U \cap (U^* \cup U^{*c})} P(x_1,...,x_n;\lambda_1) = \sum_{(x_1,..,x_n)\in U \cap U^*} P(x_1,...,x_n;\lambda_1) +$$

$$+ \sum_{(x_1,..,x_n)\in U \cap U^{*c}} P(x_1,...,x_n;\lambda_1)$$

Ţinînd seama de (7) şi (8), relaţia (6) devine

(9) $$\delta = \sum_{(x_1,..,x_n)\in U^* \cap U^c} P(x_1,...,x_n;\lambda_1) - \sum_{(x_1,...,x_n)\in U \cap U^{*c}} P(x_1,...,x_n;\lambda_1)$$

Dacă $(x_1,...,x_n) \in U^*$, atunci din ipoteză, rezultă

(10) $$P(x_1,...,x_n;\lambda_1) \geqslant kP(x_1,...,x_n;\lambda_o)$$

Intrucît $U^* \cap U^c \subset U^*$, ţinînd seama de (10), avem

(11) $$\sum_{(x_1,\dots,x_n)\in U^*\cap U^c} P(x_1,\dots,x_n;\lambda_1) \geqslant k \sum_{(x_1,\dots,x_n)\in U^*\cap U^c} P(x_1,\dots,x_n;\lambda_o)$$

Dacă $(x_1,\dots,x_n)\in U^{*c}$, atunci din ipoteză, rezultă

(12) $P(x_1,\dots,x_n;\lambda_1) \leqslant k\, P(x_1,\dots,x_n;\lambda_o)$

Intrucît $U\cap U^{*c}\subset U^{*c}$, ţinînd seama de (12), avem

(13) $$\sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_1) \leqslant k \sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_o)$$

ceea ce este echivalent cu

(14) $$-\sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_1) \geqslant -k \sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_o)$$

Ţinînd seama de (2), (3), (11) şi (14), expresia lui δ , dată de (9), devine

(15) $$\delta \geqslant k\left[\sum_{(x_1,\dots,x_n)\in U^*\cap U^c} P(x_1,\dots,x_n;\lambda_o) - \sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_o)\right] =$$

$$= k\left[\sum_{(x_1,\dots,x_n)\in U^*\cap U^c} P(x_1,\dots,x_n;\lambda_o) + \sum_{(x_1,\dots,x_n)\in U\cap U^*} P(x_1,\dots,x_n;\lambda_o) - \right.$$

$$\left. - \sum_{(x_1,\dots,x_n)\in U\cap U^*} P(x_1,\dots,x_n;\lambda_o) - \sum_{(x_1,\dots,x_n)\in U\cap U^{*c}} P(x_1,\dots,x_n;\lambda_o)\right] =$$

$$= k\left[\sum_{(x_1,\dots,x_n)\in U^*} P(x_1,\dots,x_n;\lambda_o) - \sum_{(x_1,\dots,x_n)\in U} P(x_1,\dots,x_n;\lambda_o)\right] = k(\alpha-\alpha) = 0$$

Relaţia (15) constituie ceea ce trebuia demonstrat.

§ 4. TESTE PENTRU VERIFICAREA IPOTEZEI CĂ MEDIA UNEI COLECTIVITĂŢI NORMALE ESTE EGALĂ CU O VALOARE DATĂ ÎN CAZUL CÎND DISPERSIA COLECTIVITĂŢII ESTE CUNOSCUTĂ

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd repartiţia $N(m, \sigma^2)$, unde $m = M(X)$ este parametrul necunoscut asupra căruia se face ipoteza, iar $\sigma^2 = D^2(X)$ este cunoscut. Fie m_o o valoare reală fixată. Din colectivitate se efectuează selecţia repetată $(x_1, \ldots, x_n)$. Pe baza acestei selecţii trebuie testată ipoteza că media colectivităţii este egală cu m_o. In acest scop se folosesc testele : U bilateral, U unilateral stînga şi U unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(X,n) = \frac{m(X,n) - m}{\frac{\sigma}{\sqrt{n}}}$$

Testul U bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m = m_o$$

faţă de ipoteza alternativă

$$H_1 : m = m_1 \ , \ (m_1 \neq m_o)$$

Intrucît dispersia colectivităţii este cunoscută, putem presupune că $D^2(X) = \sigma_o^2$. Spaţiul parametrilor este $\Lambda = \{(m, \sigma_o^2)\}$, adică, dreapta $\sigma^2 = \sigma_o^2$. Ipotezele H_o şi H_1 sînt ipoteze parametrice simple. Dacă ipoteza H_o se acceptă, atunci $C(X,n)$ devine

$$U(X,n) = \frac{m(X,n) - m_o}{\frac{\sigma}{\sqrt{n}}}$$

Conform teoremei 2, §7, cap.III, statistica $U(X,n)$ urmează repartiţia $N(0,1)$. Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

$$(1) \qquad P^{(n)}(\omega^{(n)} \in \Omega^{(n)} \ ; |U(X,n)| > u_\alpha) = \alpha$$

ceea ce este echivalent cu

(2) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : |U(X,n)| < u_\alpha) = 1 - \alpha$

Folosind funcţia Laplace, relaţia (2) se mai poate scrie sub forma

(3) $$2\Phi(u_\alpha) = 1 - \alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă pe baza relaţiei (3). Valoarea statisticii U(X,n) pentru selecţia efectuată este

$$U_0 = \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}}$$

Regiunea critică a testului este

$$U = \left\{ (x_1, \ldots, x_n) : \left| \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}} \right| > u_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (-\infty ; u_\alpha) \cup (u_\alpha, +\infty)$$

Testul U bilateral este :

i) dacă $U_0 \in [-u_\alpha, u_\alpha]$, ipoteza H_0 se acceptă

ii) dacă $U_0 \notin [-u_\alpha, u_\alpha]$, ipoteza H_0 se respinge

Puterea testului U bilateral pentru verificarea ipotezei H_0 faţă de ipoteza alternativă H_1 este

(4) $\pi(U, m_1) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă}) =$

$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : |U(X,n)| > u_\alpha \mid H_1 \text{ se acceptă}) =$

$= 1 - P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : |U(X,n)| < u_\alpha \mid H_1 \text{ se acceptă}) =$

$$= 1 - P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \left| \frac{m(X,n) - m_0}{\frac{\sigma}{\sqrt{n}}} \right| < u_\alpha \mid H_1 \text{ se acceptă}\right) =$$

$$= 1 - P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : m_0 - \frac{\sigma}{\sqrt{n}} u_\alpha < m(X,n) < m_0 + \frac{\sigma}{\sqrt{n}} u_\alpha \mid H_1 \text{ se acceptă}\right).$$

Dacă ipoteza H_1 se acceptă, atunci conform teoremei 1, §7, cap.III, statistica m(X,n) urmează repartiţia $N(m_1, \frac{\sigma^2}{n})$. Ţinînd seama de aceasta, relaţia (4) devine

(5) $\bar{\pi}(U,m_1)=1-\left[\Phi\left(\frac{m_o-m_1}{\sigma}\sqrt{n}+u_\alpha\right)-\Phi\left(\frac{m_o-m_1}{\sigma}\sqrt{n}-u_\alpha\right)\right]$

Testul U unilateral stînga. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m = m_o$$

față de ipoteza alternativă

$$H_1 : m = m_1 , \ (m_1 < m_o)$$

O ipoteză alternativă de forma lui H_1 se formulează, atunci cînd valorile lui m mai mari decît m_o nu conduc la respingerea lui H_o. Pentru nivelul de semnificație α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

(6) $$P^{(n)}\{\omega^{(n)} \in \Omega^{(n)} : U(X,n) < -u_\alpha\} = \alpha$$

ceea ce este echivalent cu

(7) $$F_o(-u_\alpha) = \alpha$$

unde $F_o(x)$ este funcția de repartiție N(0,1). Folosind funcția lui Laplace, relația (7) se exprimă prin

(8) $$\frac{1}{2} + \Phi(-u_\alpha) = \alpha$$

ceea ce este echivalent cu

(9) $$\frac{1}{2} - \Phi(u_\alpha) = \alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă, pe baza relației (9).

Regiunea critică a testului este

$$U_s = \left\{(x_1,\dots,x_n) : \frac{\bar{x} - m_o}{\frac{\sigma}{\sqrt{n}}} < -u_\alpha\right\}$$

iar mulțimea din R, corespunzătoare lui U_s este

$$I_s = (-\infty, -u_\alpha)$$

Testul U unilateral stînga este :

i) dacă $U_o \geq -u_\alpha$, ipoteza H_o se acceptă

ii) dacă $U_o < -u_\alpha$, ipoteza H_o se respinge

Puterea testului U unilateral stînga pentru verificarea ipotezei H_o față de ipoteza alternativă H_1 este

(5) $\bar{\pi}(U,m_1)=1-\left[\Phi\left(\frac{m_o-m_1}{\sigma}\sqrt{n}+u_\alpha\right)-\Phi\left(\frac{m_o-m_1}{\sigma}\sqrt{n}-u_\alpha\right)\right]$

Testul U unilateral stînga. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m = m_o$$

față de ipoteza alternativă

$$H_1 : m = m_1 , \ (m_1 < m_o)$$

O ipoteză alternativă de forma lui H_1 se formulează, atunci cînd valorile lui m mai mari decît m_o nu conduc la respingerea lui H_o. Pentru nivelul de semnificație α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

(6) $$p^{(n)}\{\omega^{(n)} \in \Omega^{(n)} : U(X,n) < -u_\alpha\} = \alpha$$

ceea ce este echivalent cu

(7) $$F_o(-u_\alpha) = \alpha$$

unde $F_o(x)$ este funcția de repartiție N(0,1). Folosind funcția lui Laplace, relația (7) se exprimă prin

(8) $$\frac{1}{2} + \Phi(-u_\alpha) = \alpha$$

ceea ce este echivalent cu

(9) $$\frac{1}{2} - \Phi(u_\alpha) = \alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă, pe baza relației (9).

Regiunea critică a testului este

$$U_s = \left\{(x_1,\dots,x_n) : \frac{\bar{x} - m_o}{\frac{\sigma}{\sqrt{n}}} < -u_\alpha\right\}$$

iar mulțimea din R, corespunzătoare lui U_s este

$$I_s = (-\infty, -u_\alpha)$$

Testul U unilateral stînga este :

i) dacă $U_o \geqslant -u_\alpha$, ipoteza H_o se acceptă

ii) dacă $U_o < -u_\alpha$, ipoteza H_o se respinge

Puterea testului U unilateral stînga pentru verificarea ipotezei H_o față de ipoteza alternativă H_1 este

(10) $\pi(U,m_1)=P^{(n)}(\omega^{(n)}\in\Omega(n);Y(\omega^{(n)})\in U|H_1$ se acceptă) =

$= P^{(n)}(\omega^{(n)}\in\Omega(n);U(X,n)<-u_\alpha|H_1$ se acceptă) =

$=P^{(n)}(\omega^{(n)}\in\Omega(n);\frac{m(X,n)-m_0}{\frac{\sigma}{\sqrt{n}}}<-u_\alpha|H_1$ se acceptă) =

$=P^{(n)}(\omega^{(n)}\in\Omega(n)$: $m(X,n)<m_0-\frac{\sigma}{\sqrt{n}}u_\alpha|H_1$ se acceptă) .

Dacă ipoteza H_1 se acceptă, atunci conform teoremei 1,§7, cap. III, statistica $m(X,n)$ urmează repartiţia $N(m_1,\frac{\sigma^2}{n})$. Ţinînd seama de aceasta, relaţia (10) devine

(11) $$\pi(U,m_1)=F(m_0-\frac{\sigma}{\sqrt{n}}u_\alpha)$$

unde $F(x)$ este funcţia de repartiţie $N(m_1,\frac{\sigma^2}{n})$. Folosind funţia lui Laplace, relaţia (11) devine

(12) $$\pi(U,m_1)=\frac{1}{2}+\Phi(\frac{m_0-m_1}{\sigma}\sqrt{n}-u_\alpha)$$

Testul U unilateral dreapta . Cu ajutorul acestui test se verifică ipoteza

$$H_0 : m=m_0$$

faţă de ipoteza alternativă

$$H_1 : m=m_1,\ (m_1>m_0)$$

O ipoteză alternativă de forma lui H_1 se formulează atunci cînd valorile lui m mai mici decît m_0 nu conduc la respingerea lui H_0. Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

(13) $$P^{(n)}(\omega^{(n)}\in\Omega(n) : U(X,n)>u_\alpha)=\alpha$$

ceea ce este echivalent cu

(14) $$1-F_0(u_\alpha)=\alpha$$

unde $F_0(x)$ este funcţia de repartiţie $N(0,1)$. Folosind funcţia lui Laplace, relaţia (15) se exprimă prin

(15) $$\frac{1}{2}-\Phi(u_\alpha)=\alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV di

Regiunea critică a testului este

$$U_d = \left\{ (x_1,\dots,x_n) : \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}} > u_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U_d este

$$I_d = (u_\alpha, +\infty)$$

Testul U unilateral dreapta este :

i) dacă $U_0 \leqslant u_\alpha$, ipoteza H_0 se acceptă

ii) dacă $U_0 > u_\alpha$, ipoteza H_0 se respinge

Puterea testului U unilateral dreapta pentru verificarea ipotezei H_0 faţă de ipoteza alternativă H_1 este

(16) $\bar{\pi}(U,m_1) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă}) =$

$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : U(X,n) > u_\alpha \mid H_1 \text{ se acceptă}) =$

$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : \frac{m(X,n)-m_0}{\frac{\sigma}{\sqrt{n}}} > u_\alpha \mid H_1 \text{ se acceptă}) =$

$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : m(X,n) > m_0 + \frac{\sigma}{\sqrt{n}} u_\alpha \mid H_1 \text{ se acceptă}) =$

$= 1 - P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : m(X,n) \leqslant m_0 + \frac{\sigma}{\sqrt{n}} u_\alpha \mid H_1 \text{ se acceptă})$

Dacă ipoteza H_1 se acceptă, atunci conform teoremei 1, §7, cap.III statistica $m(X,n)$ urmează repartiţia $N(m_1, \frac{\sigma^2}{n})$. Ţinînd seama de aceasta, relaţia (16) devine

(17) $$\bar{\pi}(U,m_1) = 1 - F(m_0 + \frac{\sigma}{\sqrt{n}} u_\alpha)$$

unde $F(x)$ este funcţia de repartiţie $N(m_1, \frac{\sigma^2}{n})$. Folosind funcţia lui Laplace, relaţia (17) se exprimă prin

(18) $$\bar{\pi}(U,m_1) = \frac{1}{2} + \Phi(\frac{m_0 - m_1}{\sigma} \sqrt{n} + u_\alpha) .$$

§ 5. TESTE PENTRU VERIFICAREA IPOTEZEI CĂ MEDIA UNEI COLECTIVITĂŢI NORMALE ESTE EGALĂ CU O VALOARE DATĂ ÎN CAZUL CÎND DISPERSIA COLECTIVITĂŢII ESTE NECUNOSCUTĂ.

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd repartiţia $N(m, \sigma^2)$, unde $m = M(X)$ este parametrul asupra căruia se face ipoteza, iar $\sigma^2 = D^2(X)$ este necunoscut. Fie m_o o valoare reală fixată. Din colectivitate se efectuează selecţia repetată $(x_1, \ldots, x_n)$. Pe baza acestei selecţii trebuie testată ipoteza că media colectivităţii este egală cu m_o. In acest scop se folosesc testele: t bilateral, t unilateral stînga şi t unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(X,n) = \frac{m(X,n) - m}{\sqrt{\frac{d_*^2(X,n)}{n}}}$$

Testul t bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m = m_o$$

faţă de ipoteza alternativă

$$H_1 : m = m_1, \ (m_1 \neq m_o)$$

Spaţiul parametrilor este $\Lambda = \{(m, \sigma^2)\}$. Ipotezele H_o şi H_1 sînt ipoteze parametrice compuse. Dacă ipoteza H_o se acceptă, atunci C(X,n) devine

$$t(X,n) = \frac{m(X,n) - m_o}{\sqrt{\frac{d_*^2(X,n)}{n}}}$$

Conform teoremei 9, §7, cap.III, statistica t(X,n) urmează repartiţia Student cu $n - 1$ grade de libertate. Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

$$(1) \qquad P^{(n)}\{\omega^{(n)} \in \Omega^{(n)} : |t(X,n)| > t_\alpha\} = \alpha$$

Scanned with OKEN Scanner

Valoarea critică t_α se determină cu ajutorul tabelului VII din anexă pe baza relaţiei (1). Valoarea statisticii $t(X,n)$ pentru selecţia efectuată este

$$t_o = \frac{\bar{x} - m_o}{\sqrt{\frac{s_*^2}{n}}}$$

Regiunea critică a testului este

$$U = \left\{ (x_1,\dots,x_n) : \left| \frac{\bar{x} - m_o}{\sqrt{\frac{s_*^2}{n}}} \right| > t_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (-\infty, -t_\alpha) \cup (t_\alpha, +\infty)$$

Testul t bilateral este :

i) dacă $t_o \in [-t_\alpha, t_\alpha]$, ipoteza H_o se acceptă

ii) dacă $t_o \notin [-t_\alpha, t_\alpha]$, ipoteza H_o se respinge

Puterea testului t bilateral pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

$$(2)\quad \pi(U,m_1) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă} \right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : |t(X,n)| > t_\alpha \mid H_1 \text{ se acceptă} \right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \left| \frac{m(X,n) - m_o}{\sqrt{\frac{d_*^2(X,n)}{n}}} \right| > t_\alpha \mid H_1 \text{ se acceptă} \right)$$

<u>Testul t unilateral stînga</u>. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m = m_o$$

faţă de ipoteza alternativă

$$H_1 : m = m_1 \ , \ (m_1 < m_o)$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

$$(3)\qquad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : t(X,n) < -t_\alpha \right) = \alpha$$

ceea ce este echivalent cu

(4) $$F(-t_\alpha) = \alpha$$

unde $F(x)$ este funcţia de repartiţie Student cu $n-1$ grade de libertate. Intrucît $F(-t_\alpha) = 1 - F(t_\alpha)$, valoarea critică t_α se determină cu ajutorul tabelului VI din anexă, pe baza relaţiei

(5) $$F(t_\alpha) = 1 - \alpha$$

Regiunea critică a testului este

$$U_s = \left\{(x_1,\dots,x_n) : \frac{\bar{x} - m_0}{\sqrt{\frac{s_*^2}{n}}} < -t_\alpha\right\}$$

iar mulţimea din R, corespunzătoare lui U_s este

$$I_s = (-\infty, -t_\alpha)$$

Testul t unilateral stînga este :

i) dacă $t_0 \geqslant -t_\alpha$, ipoteza H_0 se acceptă

ii) dacă $t_0 < -t_\alpha$, ipoteza H_0 se respinge

Puterea testului t unilateral stînga pentru verificarea ipotezei H_0 faţă de ipoteza alternativă H_1 este

(6) $$\pi(U,m_1) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; Y(\omega^{(n)}) \in U \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; t(X,n) < -t_\alpha \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; \frac{m(X,n)-m_0}{\sqrt{\frac{d_*^2(X,n)}{n}}} < -t_\alpha \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; m(X,n) < m_0 - t_\alpha\sqrt{\frac{d_*^2(X,n)}{n}} \mid H_1 \text{ se acceptă})$$

Testul t unilateral dreapta. Cu ajutorul acestui test se verifică ipoteza $H_0 : m = m_0$

faţă de ipoteza alternativă

$$H_1 : m = m_1, \ (m_1 > m_0)$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

(7) $$P^{(n)}(\omega^{(n)}\in\Omega^{(n)} : t(X,n) > t_\alpha) = \alpha$$

ceea ce este echivalent cu

(8) $$1 - P^{(n)}(\omega^{(n)}\in\Omega^{(n)} : t(X,n) \leq t_\alpha) = \alpha$$

Folosind funcţia de repartiţie, relaţia (8) devine

(9) $$1 - F(t_\alpha) = \alpha$$

Valoarea critică t_α se determină cu ajutorul tabelului VI din anexă, pe baza relaţiei (9).

Regiunea critică a testului este

$$U_d = \left\{(x_1,\dots,x_n) : \frac{\bar{x} - m_0}{\sqrt{\frac{s^2_*}{n}}} > t_\alpha\right\}$$

iar mulţimea din R, corespunzătoare lui U_d este

$$I_d = (t_\alpha, +\infty)$$

Testul t unilateral dreapta este :

i) dacă $t_0 \leq t_\alpha$, ipoteza H_0 se acceptă

ii) dacă $t_0 > t_\alpha$, ipoteza H_0 se respinge

Puterea testului t unilateral dreapta pentru verificarea ipotezei H_0 faţă de ipoteza alternativă H_1 este

(10) $$\bar{\pi}(U,m_1) = P^{(n)}(\omega^{(n)}\in\Omega^{(n)}; Y(\omega^{(n)})\in U \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)}\in\Omega^{(n)}; t(X,n) > t_\alpha \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)}\in\Omega^{(n)} : \frac{m(X,n)-m_0}{\sqrt{\frac{d^2_*(X,n)}{n}}} > t_\alpha \mid H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)}\in\Omega^{(n)}; m(X,n) > m_0 + t_\alpha\sqrt{\frac{d^2_*(X,n)}{n}} \mid H_1 \text{ se acceptă})$$

§ 6. TESTE PENTRU VERIFICAREA IPOTEZEI CĂ DISPERSIA UNEI COLECTIVITAŢI NORMALE ESTE EGALĂ CU O VALOARE DATĂ.

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd repartiţia $N(m,\sigma^2)$, unde $m = M(X)$, iar $\sigma^2=D^2(X)$ este parametrul asupra căruia se face ipoteza. Fie σ_o^2 o valoare reală pozitivă fixată. Din colectivitate se efectuează selecţia repetată $(x_1,\dots,x_n)$. Pe baza acestei selecţii trebuie testată ipoteza că dispersia colectivităţii este egală cu σ_o^2. In acest scop se folosesc testele: χ^2 bilateral, χ^2 unilateral stînga şi χ^2 unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(X,n) = \frac{(n-1)d_*^2(X,n)}{\sigma^2}$$

Testul χ^2 bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : \sigma^2 = \sigma_o^2$$

faţă de ipoteza alternativă

$$H_1 : \sigma^2 = \sigma_1^2 , \ (\sigma_1^2 \neq \sigma_o^2)$$

Spaţiul parametrilor este $\Lambda = \{(m,\sigma^2)\}$. Ipotezele H_o şi H_1 sînt ipoteze parametrice compuse. Dacă ipoteza H_o se acceptă, atunci $C(X,n)$ devine

$$\chi^2(X,n) = \frac{(n-1)d_*^2(X,n)}{\sigma_o^2}$$

Conform teoremei 8, §7, cap.III, statistica $\chi^2(X,n)$ urmează repartiţia hi-pătrat cu n-1 grade de libertate şi parametru $\sigma = 1$. Pentru nivelul de semnificaţie α fixat, se pot determina valorile critice a şi b, astfel încît ai aibă loc relaţiile

(1) $$P^{(n)}\{\omega^{(n)}\in\Omega^{(n)} ; a < \chi^2(X,n) < b\} = 1-\alpha$$

(2) $$P^{(n)}\{\omega^{(n)}\in\Omega^{(n)} ; \chi^2(X,n) > a\} = 1-\frac{\alpha}{2}$$

(3) $$P^{(n)}\{\omega^{(n)}\in\Omega^{(n)} ; \chi^2(X,n) > b\} = \frac{\alpha}{2} .$$

Valorile critice „a" şi „b" se determină cu ajutorul tabelului V din anexă, pe baza relaţiilor (2) şi (3). Valoarea statisticii $\chi^2(X,n)$ pentru selecţia efectuată este

$$\chi_o^2 = \frac{(n-1)s_*^2}{\sigma_o^2}$$

Regiunea critică a testului este

$$U = \left\{\left[(x_1,\dots,x_n): \frac{(n-1)s_*^2}{\sigma_o^2} \leqslant a\right] \cup \left[(x_1,\dots,x_n): \frac{(n-1)s_*^2}{\sigma_o^2} \geqslant b\right]\right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (0,a] \cup [b,+\infty)$$

Testul χ^2 bilateral este:

i) dacă $\chi_o^2 \in (a,b)$, ipoteza H_o se acceptă

ii) dacă $\chi_o^2 \notin (a,b)$, ipoteza H_o se respinge

Puterea testului χ^2 bilateral pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

$$(4)\quad \bar{\pi}(U,\sigma_1^2) = P^{(n)}\left\{\left[\omega^{(n)} \in \Omega^{(n)}: \chi^2(X,n) < a\right] \cup \left[\omega^{(n)} \in \Omega^{(n)}: \chi^2(X,n) > b\right] \mid H_1 \text{ se acceptă}\right\} =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)}: \chi^2(X,n) < a \mid H_1 \text{ se acceptă}) +$$

$$+ P^{(n)}(\omega^{(n)} \in \Omega^{(n)}: \chi^2(X,n) > b \mid H_1 \text{ se acceptă}).$$

Dacă ipoteza H_1 se acceptă, atunci statistica $\frac{(n-1)d_*^2(X,n)}{\sigma_1^2}$ urmează repartiţia hi-pătrat cu n-1 grade de libertate şi parametru $\sigma = 1$. Ţinînd seama de acest lucru relaţia (4) devine

$$(5)\quad \bar{\pi}(U,\sigma_1^2) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)}: \frac{(n-1)d_*^2(X,n)\sigma_1^2}{\sigma_o^2\sigma_1^2} < a) +$$

$$+ P^{(n)}(\omega^{(n)} \in \Omega^{(n)}: \frac{(n-1)d_*^2(X,n)\sigma_1^2}{\sigma_o^2\sigma_1^2} > b) =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)}: \frac{(n-1)d_*^2(X,n)}{\sigma_1^2} < \frac{a\sigma_o^2}{\sigma_1^2}) +$$

$$+ P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \frac{(n-1)d_*^2(X,n)}{\sigma_1^2} > \frac{b\sigma_o^2}{\sigma_1^2} \right)$$

Dacă notăm

$$\frac{(n-1)d_*^2(X,n)}{\sigma_1^2} = \tilde{\chi}^2(X,n)$$

atunci, relaţia (5) devine

$$(6) \quad \pi(U,\sigma_1^2) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \tilde{\chi}^2(X,n) < \frac{a\sigma_o^2}{\sigma_1^2} \right) +$$

$$+ P^{(n)}\left(\omega^{(n)} \in \Omega : \tilde{\chi}^2(X,n) > \frac{b\sigma_o^2}{\sigma_1^2} \right)$$

Testul χ^2 unilateral stînga. Cu ajutorul acestui test se verifică ipoteza

$$H_o : \sigma^2 = \sigma_o^2$$

faţă de ipoteza alternativă

$$H_1 : \sigma^2 = \sigma_1^2, \ (\sigma_1^2 < \sigma_o^2)$$

Pentru nivelul de semnificaţie α fixat, se determină valoarea critică „a", astfel încît să aibă loc

$$(7) \quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \chi^2(X,n) > a \right) = 1 - \alpha$$

Valoarea critică „a" se determină cu ajutorul tabelului V din anexă, pe baza relaţiei (7).

Regiunea critică a testului este

$$U_s = \left\{ (x_1,\ldots,x_n) : \frac{(n-1)s_*^2}{\sigma_o^2} < a \right\}$$

iar mulţimea din R, corepunzătoare lui U_s este

$$I_s = (0,a]$$

Testul χ^2 unilateral stînga este :

i) dacă $\chi_o^2 > a$, ipoteza H_o se acceptă

ii) dacă $\chi_o^2 < a$, ipoteza H_o se respinge .

Puterea testului χ^2 unilateral stînga pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

$$(8) \quad \pi(U,\sigma_1^2) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \frac{(n-1)d_*^2(X,n)}{\sigma_o^2} > a \mid H_1 \text{ se acceptă} \right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \tilde{\chi}^2(X,n) > \frac{a\sigma_o^2}{\sigma_1^2} \right) .$$

Testul χ^2 unilateral dreapta. Cu ajutorul acestui test se verifică ipoteza $H_o : \sigma^2 = \sigma_o^2$

față de ipoteza alternativă

$$H_1 : \sigma^2 = \sigma_1^2 (\sigma_1^2 > \sigma_o^2)$$

Pentru nivelul de semnificație α fixat se poate determina valoarea critică „b", astfel încît să aibă loc

(9) $$P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \chi^2(X,n) > b \right) = \alpha$$

Valoarea critică „b" se determină cu ajutorul tabelului V din anexă, pe baza relației (9).

Regiunea critică a testului este

$$U_d = \left\{ (x_1, \ldots, x_n) : \frac{(n-1)s_*^2}{\sigma_o^2} > b \right\}$$

iar mulțimea din R, corespunzătoare lui U_d este

$$I_d = (b, +\infty)$$

Testul χ^2 unilateral dreapta este :

i) dacă $\chi_o^2 < b$, ipoteza H_o se acceptă

ii) dacă $\chi_o^2 > b$, ipoteza H_o se respinge.

Puterea testului χ^2 unilateral dreapta pentru verificarea ipotezei H_o față de ipoteza H_1 este

(10) $$\pi(U, \sigma_1^2) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)}, \frac{(n-1)d_*^2(X,n)}{\sigma_o^2} > b \,\middle|\, H_1 \text{ se acceptă} \right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)}, \tilde{\chi}^2(X,n) > \frac{b\sigma_o^2}{\sigma_1^2} \right)$$

§ 7. TESTE PENTRU VERIFICAREA IPOTEZEI EGALITĂȚII MEDIILOR A DOUĂ COLECTIVITĂȚI NORMALE ÎN CAZUL CÎND DISPERSIILE COLECTIVITĂȚILOR SÎNT CUNOSCUTE.

Fie X_1 caracteristica urmărită la indivizii colectivității C_1 avînd repartiția $N(m_1, \sigma_1^2)$ și X_2 caracteristica urmărită la indivizii colectivității C_2 avînd repartiția $N(m_2, \sigma_2^2)$, unde $m_1 = M(X_1)$, $m_2 = M(X_2)$, iar $\sigma_1^2 = D^2(X_1)$ și $\sigma_2^2 = D^2(X_2)$ sînt cunoscuți. Din colectivitatea C_1 se efectuează selecția repetată $(x_1^{(1)}, \ldots, x_{n_1}^{(1)})$, iar din colectivitatea C_2 se efectuează selecția repetată $(x_1^{(2)}, \ldots, x_{n_2}^{(2)})$. Pe baza acestor selecții trebuie verificată ipoteza că mediile celor două colectivități sînt egale. In acest scop se folosesc testele: U bilateral, U unilateral stînga și U unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(x_1, x_2, n_1, n_2) = \frac{m(x_1, n_1) - m(x_2, n_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Testul U bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m_1 = m_2$$

față de ipoteza alternativă

$$H_1 : m_1 \neq m_2 .$$

Spațiul parametrilor este $\Lambda = \{(m_1, m_2, \sigma_1^2, \sigma_2^2)\}$. Ipotezele H_o și H_1 sînt ipoteze parametrice compuse. Dacă ipoteza H_o se acceptă, atunci $C(X,n)$ devine

$$U(X_1, X_2, n_1, n_2) = \frac{m(X_1, n_1) - m(X_2, n_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Conform teoremei 4, §7, cap.III, statistica $U(X_1, X_2, n_1, n_2)$

urmează repartiţia N(0,1). Pentru nivelul de semnificaţie α fixat se poate determina valoarea critică u_α, astfel încît să aibă loc

(1) $$P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; |U(X_1,X_2,n_1,n_2)| > u_\alpha) = \alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă, pe baza relaţiei (3) din § 4, cap.V. Valoarea statisticii $U(X_1,X_2,n_1,n_2)$ pentru cele două selecţii este

$$U_o = \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

unde

$$\overline{x^{(1)}} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i^{(1)} \quad , \quad \overline{x^{(2)}} = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i^{(2)}$$

Regiunea critică a testului este

$$U = \left\{(x_1^{(1)},\dots,x_{n_1}^{(1)}, x_1^{(2)},\dots,x_{n_2}^{(2)}) : \left| \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| > u_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (-\infty, -u_\alpha) \cup (u_\alpha, +\infty)$$

Testul U bilateral este :

i) dacă $U_o \in [-u_\alpha, u_\alpha]$, ipoteza H_o se acceptă

ii) dacă $U_o \notin [-u_\alpha, u_\alpha]$, ipoteza H_o se respinge

Notăm $m_1 - m_2 = \delta$. Dacă ipoteza H_1 se acceptă, atunci $\delta \neq 0$.

Puterea testului U bilateral pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

(2) $$\bar{\pi}(\delta) = P^{(n)}(\omega^{(n)} \in \Omega(n); \left| \frac{m(X_1,n_1) - m(X_2,n_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| > u_\alpha \mid H_1 \text{ se acceptă}) =$$

$$=P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}:\left|\frac{m(X_1,n_1)-m(X_2,n_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right|>u_\alpha \mid m_1-m_2=\delta\right)$$

Dacă ipoteza H_1 se acceptă, atunci $m(X_1,n_1)-m(X_2,n_2)$ urmează repartiţia $N(\delta,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$. Ţinînd seama de aceasta, relaţia (2) devine

$$(3)\quad \bar{\pi}(\delta)=1-P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}:\left|\frac{m(X_1,n_1)-m(X_2,n_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right|\leqslant u_\alpha \mid m_1-m_2=\delta\right)=$$

$$=1-P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)}:|m(X_1,n_1)-m(X_2,n_2)|\leqslant u_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\mid m_1-m_2=\delta\right)=$$

$$=1-\left[\Phi\left(u_\alpha-\frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right)-\Phi\left(-u_\alpha-\frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right)\right]$$

<u>Testul U unilateral stînga.</u> Cu ajutorul acestui test se verifică ipoteza

$$H_o : m_1 = m_2$$

faţă de ipoteza alternativă

$$H_1 : m_1 < m_2 .$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

$$(4)\quad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : U(X_1,X_2,n_1,n_2)<-u_\alpha\right)=\alpha .$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă, pe baza relaţiei (9) din §4, cap.V.

Regiunea critică a testului este

$$U_s=\left\{(x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{\overline{x^{(1)}}-\overline{x^{(2)}}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}<-u_\alpha\right\}$$

iar mulţimea din R, corespunzătoare lui U_s este

$$I_s = (-\infty, -u_\alpha) .$$

Testul U unilateral stînga este :

i) dacă $U_0 \geqslant -u_\alpha$, ipoteza H_0 se acceptă

ii) dacă $U_0 < -u_\alpha$, ipoteza H_0 se respinge .

Puterea testului unilateral stînga pentru verificarea ipotezei H_0 faţă de ipoteza alternativă H_1 este

$$(5)\quad \bar{\pi}(\delta) = P^{(n)}\Big(\omega^{(n)} \in \Omega^{(n)} ; \frac{m(X_1,n_1) - m(X_2,n_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < -u_\alpha \,|\, H_1 \text{ se acceptă} \Big) =$$

$$= P^{(n)}\Big(\omega^{(n)} \in \Omega^{(n)} ; m(X_1,n_1) - m(X_2,n_2) < -u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,\Big|\, m_1 - m_2 = \delta,\ \delta < 0 \Big) =$$

$$= F\Big(-u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\Big)$$

unde $F(x)$ este funcţia de repartiţie $N\big(\delta, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\big)$. Folosind funcţia lui Laplace, relaţia (5) devine

$$(6)\qquad \bar{\pi}(\delta) = \frac{1}{2} - \Phi\Big(u_\alpha - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\Big)$$

Testul U unilateral dreapta. Cu ajutorul acestui test se verifică ipoteza

$$H_0 : m_1 = m_2$$

faţă de ipoteza alternativă

$$H_1 : m_1 > m_2$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică u_α, astfel încît să aibă loc

$$(7)\qquad P^{(n)}\big(\omega^{(n)} \in \Omega^{(n)} : U(X_1, X_2, n_1, n_2) > u_\alpha \big) = \alpha$$

Valoarea critică u_α se determină cu ajutorul tabelului IV din anexă, pe baza relaţiei (15) din §4, cap.V .

Regiunea critică a testului este

$$U_d = \left\{ (x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > u_\alpha \right\}$$

Testul U unilateral dreapta este :

i) dacă $U_0 \leqslant u_\alpha$, ipoteza H_0 se acceptă

ii) dacă $U_0 > u_\alpha$, ipoteza H_0 se respinge

Puterea testului unilateral dreapta pentru verificarea ipotezei H_0 față de ipoteza H_1 este

$$(8)\quad \bar{\pi}(\delta) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \frac{m(X_1,n_1) - m(X_2,n_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > u_\alpha \,|\, H_1 \text{ se acceptă}\right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : m(X_1,n_1) - m(X_2,n_2) > u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,|\, m_1 - m_2 = \delta, \delta > 0\right) =$$

$$= 1 - P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : m(X_1,n_1) - m(X_2,n_2) \leqslant u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,|\, m_1 - m_2 = \delta, \delta > 0\right) =$$

$$= 1 - F\left(u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \frac{1}{2} - \Psi\left(u_\alpha - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$$

Observație. Testele prezentate în acest paragraf pot fi folosite chiar dacă colectivitățile C_1 și C_2 nu sînt normale, întrucît pentru n_1 și n_2 mari, conform teoremei lui Leapunov, §1, cap.I, statistica $U(X_1,X_2,n_1,n_2)$ urmează o repartiție N(0,1), dacă ipoteza H_0 se acceptă.

§ 8. TESTE PENTRU VERIFICAREA IPOTEZEI EGALITĂȚII MEDIILOR A DOUĂ COLECTIVITĂȚI NORMALE ÎN CAZUL CÎND DISPERSIILE COLECTIVITĂȚILOR SINT EGALE CU O ACEEAȘI VALOARE NECUNOSCUTĂ.

Fie X_1 caracteristica urmărită la indivizii colectivității C_1 avînd repartiția $N(m_1, \sigma^2)$ și X_2 caracteristica urmărită la indivizii colectivității C_2 avînd repartiția $N(m_2, \sigma^2)$, unde

$m_1 = M(X_1)$, $m_2 = M(X_2)$, iar $\sigma^2 = D^2(X_1) = D^2(X_2)$ este necunoscut. Din colectivitatea C_1 se efectuează selecţia repetată $(x_1^{(1)},\dots,x_{n_1}^{(1)})$, iar din colectivitatea C_2 se efectuează selecţia repetată $(x_1^{(2)},\dots,x_{n_2}^{(2)})$. Pe baza acestor selecţii trebuie verificată ipoteza că mediile celor două colectivităţi sînt egale. In acest scop se folosesc testele : t bilateral, t unilateral stînga şi t unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(X_1,X_2,n_1,n_2)=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\;\frac{m(X_1,n_1)-m(X_2,n_2)-(m_1-m_2)}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}}$$

Testul t bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m_1 = m_2$$

faţă de ipoteza alternativă

$$H_1 : m_1 \neq m_2$$

Spaţiul parametrilor este $\Lambda = \{(m_1,m_2,\sigma^2)\}$. Ipotezele H_o şi H_1 sînt ipoteze parametrice compuse. Dacă ipoteza H_o se acceptă, atunci $C(X_1,X_2,n_1,n_2)$ devine

$$t(X_1,X_2,n_1,n_2)=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\;\frac{m(X_1,n_1)-m(X_2,n_2)}{\sqrt{(n_1-1)d_*^2(X_1,n_1)+(n_2-1)d_*^2(X_2,n_2)}}$$

Conform teoremei 10, §7, cap.III, statistica $t(X_1,X_2,n_1,n_2)$ urmează repartiţia Student cu n_1+n_2-2 grade de libertate. Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

$$(1)\qquad P^{(n)}\{\omega^{(n)}\in\Omega(n) : |t(X_1,X_2,n_1,n_2)| > t_\alpha\} = \alpha$$

Valoarea critică t_α se determină cu ajutorul tabelului VII din anexă pe baza relaţiei (1). Valoarea statisticii $t(X_1,X_2,n_1,n_2)$ pentru selecţiile efectuate este

$$t_o = \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \; \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{(n_1-1)s_*^{(1)2}+(n_2-1)s_*^{(2)2}}}$$

Regiunea critică a testului este

$$U = \left\{ (x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \left| \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \; \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{(n_1-1)s_*^{(1)2}+(n_2-1)s_*^{(2)2}}} \right| > t_\alpha \right\}$$

Testul t bilateral este :

i) dacă $t_o \in [-t_\alpha, t_\alpha]$, ipoteza H_o se acceptă

ii) dacă $t_o \notin [-t_\alpha, t_\alpha]$, ipoteza H_o se respinge

Fie $\delta = m_1 - m_2$. Puterea testului t bilateral pentru verificarea ipotezei H_o față de ipoteza alternativă H_1 este

$$(2)\quad \pi(\delta) = P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : |t(X_1,X_2,n_1,n_2)| > t_\alpha \mid m_1-m_2=\delta \right)$$

<u>Testul t unilateral stînga</u>. Cu ajutorul acestui test se verifică ipoteza

$$H_o : m_1 = m_2$$

față de ipoteza alternativă

$$H_1 : m_1 < m_2 .$$

Pentru nivelul de semnificație α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

$$(3)\qquad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : t(X_1,X_2,n_1,n_2) < -t_\alpha \right) = \alpha$$

Valoarea critică t_α se determină cu ajutorul tabelului VI pe baza relației (5) din §5, cap.V.

Regiunea critică a testului este

$$U_s = \left\{ (x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \; \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{(n_1-1)s_*^{(1)2}+(n_2-1)s_*^{(2)2}}} < -t_\alpha \right\}$$

iar mulțimea din R, corespunzătoare lui U_s este

$$I_s = (-\infty, -t_\alpha)$$

Testul t unilateral stînga este

i) dacă $t_o \geqslant -t_\alpha$, ipoteza H_o se acceptă

ii) dacă $t_o < -t_\alpha$, ipoteza H_o se respinge

Puterea testului t unilateral stînga pentru verificarea ipotezei H_o față de ipoteza alternativă H_1 este

(4) $\bar{\pi}(\delta) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; t(X_1, X_2, n_1, n_2) < -t_\alpha \mid m_1 - m_2 = \delta\,,\ \delta < 0)$

Testul t unilateral dreapta. Cu ajutorul acestui test se verifică ipoteza $H_o : m_1 = m_2$

față de ipoteza alternativă

$$H_1 : m_1 > m_2$$

Pentru nivelul de semnificație α fixat, se poate determina valoarea critică t_α, astfel încît să aibă loc

(5) $P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; t(X_1, X_2, n_1, n_2) > t_\alpha) = \alpha$

Valoarea critică t_α se determină cu ajutorul tabelului VI din anexă, pe baza relației (9) din §5, cap.V.

Regiunea critică a testului este

$$U_d = \left\{ (x_1^{(1)}, \ldots, x_{n_1}^{(1)}, x_1^{(2)}, \ldots, x_{n_2}^{(2)}) : \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \, \frac{\overline{x^{(1)}} - \overline{x^{(2)}}}{\sqrt{(n_1-1)s_*^{(1)2} + (n_2-1)s_*^{(2)2}}} > t_\alpha \right\}$$

iar mulțimea din R, corepunzătoare lui U_d este

$$I_d = (t_\alpha, +\infty)\,.$$

Testul t unilateral dreapta este :

i) dacă $t_o \leqslant t_\alpha$, ipoteza H_o se acceptă

ii) dacă $t_o > t_\alpha$, ipoteza H_o se respinge

Puterea testului t unilateral stînga pentru verificarea ipotezei față de ipoteza H_1 este

(6) $\bar{\pi}(\delta) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} ; t(X_1, X_2, n_1, n_2) > t_\alpha \mid m_1 - m_2 = \delta\,,\ \delta > 0).$

§ 9. TESTE PENTRU VERIFICAREA IPOTEZEI EGALITĂŢII DISPERSIILOR A DOUĂ COLECTIVITĂŢI NORMALE.

Fie X_1 caracteristica urmărită la indivizii colectivităţii C_1 avînd repartiţia $N(m_1, \sigma_1^2)$ şi X_2 caracteristica urmărită la indivizii colectivităţii C_2 avînd repartiţia $N(m_2, \sigma_2^2)$, unde $m_1 = M(X_1), m_2 = M(X_2), \sigma_1^2 = D^2(X_1), \sigma_2^2 = D^2(X_2)$. Din colectivitatea C_1 se efectuează selecţia repetată $(x_1^{(1)}, \ldots, x_{n_1}^{(1)})$ iar din colectivitatea C_2 se efectuează selecţia repetată $(x_1^{(2)}, \ldots, x_{n_2}^{(2)})$. Pe baza acestor selecţii trebuie verificată ipoteza că dispersiile celor două colectivităţi sînt egale. In acest scop se folosesc testele : F bilateral, F unilateral stînga şi F unilateral dreapta. Toate aceste teste se bazează pe criteriul

$$C(X_1, X_2, n_1, n_2) = \frac{\sigma_2^2 d_*^2(X_1, n_1)}{\sigma_1^2 d_*^2(X_2, n_2)} .$$

Testul F bilateral. Cu ajutorul acestui test se verifică ipoteza

$$H_o : \sigma_1^2 = \sigma_2^2$$

faţă de ipoteza alternativă

$$H_1 : \sigma_1^2 \neq \sigma_2^2 .$$

Spaţiul parametrilor este $\Lambda = \{(m_1, m_2, \sigma_1^2, \sigma_2^2)\}$. Ipotezele H_o şi H_1 sînt ipoteze parametrice compuse. Dacă ipoteza H_o se acceptă, atunci $C(X_1, X_2, n_1, n_2)$ devine

$$F(X_1, X_2, n_1, n_2) = \frac{d_*^2(X_1, n_1)}{d_*^2(X_2, n_2)} .$$

Conform teoremei 11, §7, cap.III, statistica $F(X_1, X_2, n_1, n_2)$ urmează repartiţia Snedecor cu n_1-1 şi n_2-1 grade de libertate. Pentru nivelul de semnificaţie α fixat, se pot determi

valorile critice „a" şi „b" , astfel încît să aibă loc relaţiile

(1) $$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : a \leqslant F(X_1,X_2,n_1,n_2) \leqslant b\right) = 1-\alpha$$

(2) $$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : F(X_1,X_2,n_1,n_2) > a\right) = 1-\frac{\alpha}{2}$$

(3) $$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : F(X_1,X_2,n_1,n_2) > b\right) = \frac{\alpha}{2}$$

Valorile critice „a" şi „b"se determină cu ajutorul tabelelor VIII sau IX din anexă, pe baza relaţiilor (2) şi (3).

Observaţie. Dacă statistica $F(X_1,X_2,n_1,n_2)$ urmează o repartiţie Snedecor cu n_1-1 şi n_2-1 grade de libertate, atunci statistica

$$S(X_1,X_2,n_1,n_2) = \frac{1}{F(X_1,X_2,n_1,n_2)}$$

urmează repartiţia Snedecor cu n_2-1 şi n_1-1 grade de libertate.

Ţinînd seama de observaţia făcută şi relaţia (2), valoarea critică „a" se mai poate determina pe baza relaţiei

(4) $$P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : F(X_1,X_2,n_1,n_2) \leqslant a\right) =$$

$$= P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : \frac{1}{F(X_1,X_2,n_1,n_2)} \geqslant \frac{1}{a}\right) =$$

$$= P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)} : S(X_1,X_2,n_1,n_2) \geqslant \frac{1}{a}\right) = \frac{\alpha}{2} .$$

Valoarea statisticii $F(X_1,X_2,n_1,n_2)$ pentru selecţiile efectuate este

$$F_o = \frac{s_*^{(1)2}}{s_*^{(2)2}} .$$

Regiunea critică a testului este

$$U = \left\{\left[(x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{s_*^{(1)2}}{s_*^{(2)2}} \leqslant a\right] \cup \right.$$

$$\left. \cup\left[(x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{s_*^{(1)2}}{s_*^{(2)2}} \geqslant b\right]\right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (0,a] \cup [b,+\infty)$$

Testul F bilateral este :

i) dacă $F_o \in (a,b)$, ipoteza H_o se acceptă

ii) dacă $F_o \notin (a,b)$, ipoteza H_o se respinge.

Notăm $\frac{\sigma_1^2}{\sigma_2^2} = \lambda$. Dacă ipoteta H_1 se acceptă, atunci $\lambda \neq 1$.

Puterea testului F bilateral pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

$$(5) \quad \overline{\pi}(\lambda) = P^{(n)}\left[\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) < a\right) \cup \left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > b\right) | H_1 \text{ se acceptă}\right] =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) < a | H_1 \text{ se acceptă}\right) +$$

$$+ P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > b | H_1 \text{ se acceptă}\right).$$

Dacă ipoteza H_1 se acceptă, atunci relaţia (5) devine

$$(6) \quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \frac{\sigma_2^2 d_*^2(X_1,n_1)\sigma_1^2}{\sigma_1^2 d_*^2(X_2,n_2)\sigma_2^2} < a\right) +$$

$$+ P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : \frac{\sigma_2^2 d_*^2(X_1,n_1)\sigma_1^2}{\sigma_1^2 d_*^2(X_2,n_2)\sigma_2^2} > b\right) =$$

$$= P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) < \frac{a}{\lambda}\right) +$$

$$+ P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > \frac{b}{\lambda}\right)$$

<u>Testul F unilateral stînga</u>. Cu ajutorul acestui test se verifică ipoteza

$$H_o : \sigma_1^2 = \sigma_2^2$$

faţă de ipoteza alternativă

$$H_1 : \sigma_1^2 < \sigma_2^2 .$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică „a" , astfel încît să aibă loc

$$(7) \quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) < a\right) = \alpha$$

ceea ce este echivalent cu

$$(8) \quad P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > a\right) = 1 - \alpha$$

Valoarea critică „a" se determină cu ajutorul tabelelor VIII sau IX din anexă, pe baza relaţiei (8).

Regiunea critică a testului este

$$U_s = \left\{(x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{s_*^{(1)2}}{s_*^{(2)2}} < a\right\}$$

iar mulţimea din R, corespunzătoare lui U_s este

$$I_s = (0,a)$$

Testul F unilateral stînga este :

i) dacă $F_o \geqslant a$, ipoteza H_o se acceptă

ii) dacă $F_o < a$, ipoteza H_o se respinge

Puterea testului F unilateral stînga pentru verificarea ipotezei H_o faţă de ipoteza alternativă H_1 este

$$(9)\quad \bar{\pi}(\lambda)=P^{(n)}(\omega^{(n)}\in\Omega^{(n)}; F(X_1,X_2,n_1,n_2) < a | H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)}\in\Omega^{(n)} : F(X_1,X_2,n_1,n_2) < \frac{a}{\lambda})$$

<u>Testul F unilateral dreapta</u>. Cu ajutorul acestui test se verifică ipoteza

$$H_o : \sigma_1^2 = \sigma_2^2$$

faţă de ipoteza alternativă

$$H_1 : \sigma_1^2 > \sigma_2^2 .$$

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică b, astfel încît să aibă loc

$$(10)\qquad P^{(n)}(\omega^{(n)}\in\Omega^{(n)} : F(X_1,X_2,n_1,n_2) > b) = \alpha$$

Valoarea critică „b" se determină cu ajutorul tabelelor VIII sau IX din anexă, pe baza relaţiei (10).

Regiunea critică a testului este

$$U_d = \left\{(x_1^{(1)},\dots,x_{n_1}^{(1)},x_1^{(2)},\dots,x_{n_2}^{(2)}) : \frac{s_*^{(1)2}}{s_*^{(2)2}} > b\right\}$$

iar mulţimea din R, corespunzătoare lui U_d este

$$I_d = (b,+\infty) .$$

Testul F unilateral dreapta este :

i) dacă $F_o \leqslant b$, ipoteza H_o se acceptă

ii) dacă $F_o > b$, ipoteza H_o se respinge .

Puterea testului F unilateral dreapta pentru verificarea ipotezei H_o față de ipoteza alternativă H_1 este

$$(11)\quad \pi(\lambda) = P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > b \,|\, H_1 \text{ se acceptă}) =$$

$$= P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : F(X_1,X_2,n_1,n_2) > \frac{b}{\lambda})$$

§ 10. TESTUL RAPORTULUI DE VEROSIMILITĂȚI.

Fie X o caracteristică avînd densitatea de repartiție $f(x;\lambda_1,\ldots,\lambda_k)$ unde $(\lambda_1,\ldots,\lambda_k) \in \Lambda \subseteq R^k$. Fie B o mulțime boreliană a lui Λ . Din colectivitatea la indivizii căreia se urmărește caracteristica X se efectuează selecția repetată $(x_1,\ldots,x_n)$. Pe baza acestei selecții trebuie verificată ipoteza

$$H : (\lambda_1,\ldots,\lambda_k) \in B$$

față de ipoteza alternativă

$$H' : (\lambda_1,\ldots,\lambda_k) \notin B .$$

Fie $\hat{\lambda}_1(X,n),\ldots,\hat{\lambda}_k(X,n)$ funcții de estimație de verosimilitate maximă, respectiv pentru $\lambda_1,\ldots,\lambda_k$. Dacă $V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_k)$ este verosimilitatea, atunci $V[x_1,\ldots,x_n;\hat{\lambda}_1(X,n),\ldots,\hat{\lambda}_k(X,n)]$ va fi o statistică, pe care o vom nota $V(X,n)$. Prin urmare,

$$V(X,n) = \sup_{(\lambda_1,\ldots,\lambda_k)} V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_k)$$

Fie

$$(1)\quad V_B(X,n) = \sup_{(x_1,\ldots,x_n)\in B} V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_k)$$

$$(2)\quad V_\Lambda(X,n) = \sup_{(x_1,\ldots,x_n)\in \Lambda} V(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_k)$$

Din $B \subseteq \Lambda$, rezultă

$$(3)\quad V_B(X,n) \leqslant V_\Lambda(X,n)$$

oricare ar fi vectorul de selecţie $(X_1,\dots,X_n)$ al caracteristicii X.

Definiţie. Se numeşte raport de verosimilităţi pentru verificarea ipotezei H faţă de ipoteza alternativă H', statistica

$$L(X,n) = \frac{V_B(X,n)}{V_\Lambda(X,n)} .$$

Din definiţia raportului de verosimilităţi şi relaţia (3), rezultă

$$0 < L(X,n) \leqslant 1 \tag{4}$$

oricare ar fi vectorul de selecţie $(X_1,\dots,X_n)$ al caracteristicii X.

Presupunem că dacă ipoteza H se acceptă, statistica $L(X,n)$ are densitatea de repartiţie complet specificată $g(\ell)$. Pentru nivelul de semnificaţie α fixat, se poate determina valoarea critică L_α, astfel în cît să aibă loc

$$P^{(n)}\{\omega^{(n)} \in \Omega^{(n)} : L(X,n) < L_\alpha\} = \alpha \tag{5}$$

ceea ce este echivalent cu

$$\int_0^{L_\alpha} g(\ell)\,d\ell = \alpha \tag{6}$$

Valoarea statisticii $L(X,n)$ pentru selecţia efectuată este

$$L_0 = \frac{\sup\limits_{(\lambda_1,\dots,\lambda_k)\in B} V(x_1,\dots,x_n;\lambda_1,\dots,\lambda_k)}{\sup\limits_{(\lambda_1,\dots,\lambda_k)\in \Lambda} V(x_1,\dots,x_n;\lambda_1,\dots,\lambda_k)} = \frac{V_B(x_1,\dots,x_n;\bar\lambda_1,\dots,\bar\lambda_k)}{V_\Lambda(x_1,\dots,x_n;\bar\lambda_1,\dots,\bar\lambda_k)}$$

unde $\bar\lambda_1,\dots,\bar\lambda_k$ sînt estimaţii de verosimilitate maximă, respectiv pentru parametrii $\lambda_1,\dots,\lambda_k$.

Regiunea critică a testului este

$$U = \{(x_1,\dots,x_n) : L_0 < L_\alpha\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (0, L_\alpha) .$$

Testul raportului de verosimilităţi este :

i) dacă $L_0 \geq L_\alpha$, ipoteza H se acceptă

ii) dacă $L_0 < L_\alpha$, ipoteza H se respinge

In cazul selecţiilor mari, are loc

Teorema 1. Dacă spaţiul parametrilor Λ este k-dimensional, mulţimea boreliană B este s-dimensională ($s < k$) şi ipoteza H se acceptă, atunci statistica

$$-2 \ell_n L(X,n)$$

urmează asimptotic repartiţia hi-pătrat cu k-s grade de libertate şi parametru $\sigma = 1$.

Demonstraţie. A se vedea Ciucu G., Craiu V. [2].

§ 11. TESTE CU PRIVIRE LA FUNCŢIA DE REPARTIŢIE A COLECTIVITĂŢII.

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd funcţia de repartiţie F(x). In cele mai multe cazuri, funcţia de repartiţie a colectivităţii este complet necunoscută. Reprezentarea grafică a repartiţiilor statistice ale colectivităţii sugerează repartiţia probabilistică pe care o are X. Astfel se ajunge la formularea ipotezei statistice

H : Funcţia de repartiţie a colectivităţii este $F_0(x)$.

Funcţia de repartiţie $F_0(x)$, numită funcţie de repartiţie ipotetică, poate fi numai specificată sau complet specificată. Ipoteza H este o ipoteză neparametrică. Verificarea ipotezei H se face pe baza unei selecţii repetate de volum n. Presupunem că selecţiei efectuate îi corespunde repartiţia statistică

$$\begin{bmatrix} x_i \\ n_i^o \end{bmatrix}_{1 \leq i \leq k} , \quad \sum_{i=1}^{k} n_i^o = n \ .$$

In acest paragraf vom prezenta testul lui K.A. Pearson şi A. N. Kolmogorov pentru testarea ipotezei H.

Testul lui K.A. Pearson. Acest test este cunoscut în literatura de specialitate sub numele de testul hi-pătrat . Notăm $L = X(\Omega)$. Fie L_i, $1 \leq i \leq k$, mulţimi boreliene ale lui R, ce realizează o partiţie a lui L. Prin urmare, au loc relaţiile

$$L = \bigcup_{i=1}^{k} L_i \ , \ L_i \cap L_j = \emptyset \ , \ i \neq j \ , \ 1 \leq i,j \leq k \ .$$

Se consideră evenimentele

$$A_i = \{\omega^{(n)} \in \Omega^{(n)} \ ; \ X_i(\omega^{(n)}) \in L_i\} \ , \ 1 \leq i \leq k \ .$$

Ţinînd seama de definiţia 3, §1, cap.III, evenimentele A_i se mai pot exprima prin

$$A_i = \{\omega^{(n)} \in \Omega^{(n)} \ ; \ X(\omega_i) \in L_i\} \ , \ 1 \leq i \leq k \ .$$

Pentru i fixat, A_i reprezintă evenimentul pentru care caracteristica X ia o valoare din mulţimea boreliană L_i. Prin urmare, din cele n valori obţinute în selecţie, numai k valori sînt distincte şi ele se repartizează în k mulţimi boreliene ale lui R. Notăm $P^{(n)}(A_i) = p_i$, $1 \leq i \leq k$. Evident, au loc relaţiile

$$\Omega^{(n)} = \bigcup_{i=1}^{k} A_i \ , \ A_i \cap A_j = \emptyset \ , \ i \neq j \ , \ 1 \leq i,j \leq k$$

de unde rezultă că $\sum_{i=1}^{k} p_i = 1$. Fie numerele $p_i^o \geq 0$, $1 \leq i \leq k$, ce satisfac relaţia $\sum_{i=1}^{k} p_i^o = 1$. Ipoteza H poate fi înlocuită cu ipoteza echivalentă

$$H_1 \ : \ p_i = p_i^o \ , \ 1 \leq i \leq k$$

unde p_i^o se calculează cu ajutorul lui $F_o(x)$.

Observaţie. Dacă $F_o(x)$ este numai specificată, atunci înainte de a verifica ipoteza se estimează parametri necunoscuţi ce intră în $F_o(x)$. Estimarea parametrilor se poate face fie prin metoda modificată a minimului lui hi-pătrat, fie prin oricare altă metodă de estimare din §6, cap.IV.

Pentru i fixat, $n_i(\omega^{(n)})$ este frecvenţa absolută a evenimentului A_i în orice selecţie repetată de volum n. Mărimea $n_i(\omega^{(n)})$ este o variabilă aleatoare definită pe $\Omega^{(n)}$, cu valorile: $0,1,2,\ldots,n$, urmînd o repartiţie binominală. Dacă n_i^o este valoarea lui $n_i(\omega^{(n)})$ pentru $\omega^{(n)}$ fixat, $1 \leq i \leq k$, atunci evident $\sum_{i=1}^{k} n_i^o = n$. Testul lui K.A. Pearson se bazează pe criteriul

$$C(X,n) = \sum_{i=1}^{k} \frac{[n_i(\omega^{(n)}) - np_i]^2}{np_i} .$$

Mărimea $C(X,n)$ este o statistică prin intermediul variabilelor aleatoare $n_i(\omega^{(n)})$, $1 \leq i \leq k$. Dacă ipoteza H_1 se acceptă, criteriul $C(X,n)$ devine

$$\chi^2(X,n) = \sum_{i=1}^{k} \frac{[n_i(\omega^{(n)}) - np_i^o]^2}{np_i^o} .$$

Se poate demonstra (a se vedea Corina Reischer, Sâmboan G. Theodorescu R. [15]) că, pentru $n \to \infty$, statistica $\chi^2(X,n)$ urmează o repartiţie hi-pătrat de parametru $\sigma = 1$ şi

i) $k - 1$ grade de libertate, dacă $F_o(x)$ este complet specificată.

ii) $k-r-1$ grade de libertate dacă $F_o(x)$ conţine r parametri necunoscuţi.

Pentru nivelul de semnificaţie α fixat, cu ajutorul tabelului V din anexă, se determină valoarea critică χ^2_α, astfel încît să aibă loc

$$P^{(n)}(\omega^{(n)} \in \Omega^{(n)} : \chi^2(X,n) > \chi^2_\alpha) = \alpha .$$

Valoarea statisticii $\chi^2(X,n)$ pentru selecţia efectuată este

$$\chi^2_o = \sum_{i=1}^{k} \frac{(n_i^o - np_i^o)^2}{np_i^o} .$$

Regiunea critică a testului este

$$U = \left\{ (x_1,\ldots,x_n) : \sum_{i=1}^{k} \frac{(n_i^o - np_i^o)^2}{np_i^o} > \chi^2_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (\chi^2_\alpha, +\infty) .$$

Testul hi-pătrat este :

i) dacă $\chi^2_o < \chi^2_\alpha$, ipoteza H_1 se acceptă

ii) dacă $\chi^2_o > \chi^2_\alpha$, ipoteza H_1 se respinge .

Testul lui A.N. Kolmogorov. Cu ajutorul acestui test se verifică ipoteza H, numai în cazul cînd funcţia de repartiţie ipotetică $F_o(x)$ este continuă. Acest test se bazează pe teorema 5, §2, cap.III. Drept criteriu de verificare se foloseşte

$$C(X,n) = \sup_{x \in R} |F^*_n(x, \omega^{(n)}) - F(x)| .$$

Pentru x fixat, mărimea C(X,n) este o statistică prin intermediul lui $F^*_n(x, \omega^{(n)})$. Dacă ipoteza H se acceptă, criteriul C(X,n) devine

$$D(X,n) = \sup_{x \in R} |F^*_n(x, \omega^{(n)}) - F_o(x)| .$$

Pentru nivelul de semnificaţie α fixat, cu ajutorul tabelului X din anexă, se determină valoarea λ_α, încît să aibă loc

$$K(\lambda_\alpha) = 1 - \alpha$$

unde $K(\lambda)$ este funcţia lui Kolmogorov. Valoarea statisticii D(X,n) pentru selecţia efectuată este

$$D_o = \max_{1 \le i \le n} \{|\bar{F}_n(x_i) - F_o(x_i)|\}$$

unde $\bar{F}_n(x)$ este funcţia de frecvenţă cumulată .

Regiunea critică a testului este

$$U = \left\{(x_1, \ldots, x_n) : D_o \ge \frac{\lambda_\alpha}{\sqrt{n}}\right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = \left[\frac{\lambda_\alpha}{\sqrt{n}}, +\infty\right) .$$

Testul lui Kolmogorov este :

i) dacă $D_o < \frac{\lambda_\alpha}{\sqrt{n}}$, ipoteza H se acceptă

ii) dacă $D_o > \frac{\lambda_\alpha}{\sqrt{n}}$, ipoteza H se respinge .

§ 12. TESTUL SECVENȚIAL AL RAPORTULUI PROBABILITĂȚILOR.

In toate problemele considerate pînă acum, unde s-a folosit metoda selecției, volumul selecției era fixat dinainte. In cazul analizei secvențiale, datorită lui A.Wald, volumul selecției este o variabilă aleatoare. La verificarea unei ipoteze statistice, cu ajutorul unui test secvențial, se ia una din deciziile:

i) acceptarea ipotezei
ii) respingerea ipotezei
iii) continuarea procesului de selecție prin cercetarea a încă unei unități statistice.

Fie X caracteristica urmărită la indivizii unei colectivități, avînd densitatea de repartiție $f(x;\lambda)$, unde λ poate lua numai valorile λ_o și λ_1 . Trebuie verificată ipoteza

$$H_o : \lambda = \lambda_o$$

față de ipoteza alternativă

$$H_1 : \lambda = \lambda_1$$

cu ajutorul testului secvențial al raportului probabilităților. Din colectivitate se efectuează selecții și se consideră raportul

$$\frac{\prod_{i=1}^{\nu} f(x_i;\lambda_1)}{\prod_{i=1}^{\nu} f(x_i;\lambda_o)}$$

unde ν este o variabilă aleatoare ce poate lua valorile 1,2,3,... Notăm

$$\frac{p_{1\nu}}{p_{o\nu}} = \frac{\prod_{i=1}^{\nu} f(x_i;\lambda_1)}{\prod_{i=1}^{\nu} f(x_i;\lambda_o)} .$$

Dacă se fixează riscurile α și β , atunci se pot determina în mod unic, numerele A și B, astfel încît :

i) dacă $\frac{p_{1\nu}}{p_{0\nu}} \leqslant A$, se acceptă ipoteza H_0

ii) dacă $\frac{p_{1\nu}}{p_{0\nu}} \geqslant B$, se acceptă ipoteza H_1

iii) dacă $A < \frac{p_{1\nu}}{p_{0\nu}} < B$, se continuă selecția efectuînd observația de rang $\nu + 1$.

Valorile A și B se aproximează prin :

$$A \simeq \frac{\beta}{1-\alpha} \quad , \quad B \simeq \frac{1-\beta}{\alpha} \quad .$$

P R O B L E M E

1. Conținutul de crom dintr-un anumit material este o caracteristică ce urmează repartiția normală cu abaterea medie pătratică $\sigma = 0{,}23$. Să se testeze, la nivelul de semnificație $\alpha = 0{,}02$, ipoteza că, conținutul mediu de crom este de 29 %, folosind selecția

$x_1 = 28{,}35$ $x_2 = 28{,}80$ $x_3 = 28{,}50$ $x_4 = 28{,}75$ $x_5 = 28{,}59$
$x_6 = 28{,}99$ $x_7 = 28{,}70$ $x_8 = 28{,}38$ $x_9 = 28{,}52$.

2. Durata de ardere a unui tub fluorescent este o caracteristică ce urmează repartiția normală. Să se testeze, la nivelul de semnificație $\alpha = 0{,}02$, ipoteza că, durata medie de ardere este de 2000 ore, folosind selecția

$x_1 = 1960$ $x_2 = 1980$ $x_3 = 2000$ $x_4 = 2020$ $x_5 = 2040$.

3. Greutatea firelor de lînă este o caracteristică ce urmează repartiția normală. Să se testeze, la nivelul de semnificație $\alpha = 0{,}02$ ipoteza că, dispersia greutății firului de lînă este egală cu 0,073, folosind selecția

$x_1=7,50$ $x_2=8,22$ $x_3=7,78$ $x_4=7,70$ $x_5=8,24$ $x_6=8,09$
$x_7=7,96$ $x_8=8,12$ $x_9=7,96$ $x_{10}=8,24$ $x_{11}=8,05$ $x_{12}=7,91$.

4. Numărul metric al firelor produse de carde este o caracteristică ce urmează repartiţia normală. Dispersiile numărului metric al firelor produse la două carde sînt respectiv, $\sigma_1^2 = 0,005$ şi $\sigma_2^2 = 0,004$. Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza că, numărul metric mediu al firelor produse la cele două carde este acelaşi, folosind datele din tabelul 4

T a b e l u l 4

Numărul metric al firelor de la	I-a cardă	0,245	0,240	0,235	0,250	0,246	0,255
	II-a cardă	0,220	0,228	0,225	0,230	0,235	0,245

5. Conţinutul de siliciu al şarjelor de oţel este o caracteristică ce urmează repartiţia normală. Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza că, conţinutul mediu de siliciu determinat prin metoda chimică şi metoda calorimetrică este acelaşi, folosind datele din tabelul 5

T a b e l u l 5

Metoda chimică	1,28	0,98	1,30	1,55	1,08	1,37	1,21
Metoda calori-metrică	1,33	1,01	1,29	1,53	1,15	1,38	1,18

6. Producţia de grîu obţinută de pe două loturi de pămînt este o caracteristică ce urmează repartiţia normală. Fiecare din cele două loturi este împărţit în 10 parcele. Unul din loturi se tratează cu îngrăşăminte ce conţin fosfor. Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza egalităţii

dispersiilor producţiilor de grîu obţinute de pe cele două loturi, folosind datele din tabelul 6

T a b e l u l 6

Producţia de pe lotul tratat	6,2	5,7	6,3	6,0	6,4	5,8	5,5	6,1	5,9	6,5
Producţia de pe lotul netratat	5,6	5,9	5,6	5,7	5,5	5,4	5,2	5,8	5,3	5,1

7. Caracteristica urmărită la indivizii unei colectivităţi are repartiţia normală de medie m necunoscută şi dispersie $\sigma^2 = 1$. Din colectivitate se efectuează selecţia repetată $(x_1,\dots,x_i,\dots,x_n)$. Să se testeze, la nivelul de semnificaţie $\alpha = 0{,}05$, ipoteza $H_o : m = 3$ faţă de ipoteza alternativă $H_1 : m \neq 3$ folosind testul raportului de verosimilităţi.

8. Caracteristica unei colectivităţi urmează repartiţia normală de medie m necunoscută şi dispersie $\sigma^2 = 1$. Din colectivitate se efectuează selecţia repetată $(x_1,\dots,x_i,\dots,x_n)$. Să se testeze ipoteza $H_o : m = m_o$ faţă de ipoteza alternativă $H_1 : m = m_1$ folosind testul secvenţial al raportului probabilităţilor cu α şi β daţi .

9. Caracteristica X a unei colectivităţi urmează o repartiţie probabilistică dată de

$$X : \begin{bmatrix} 0 & 1 \\ q & p \end{bmatrix}, \quad p + q = 1 .$$

Din colectivitate se efectuează o selecţie repetată de volum n. Să se testeze ipoteza $H_o : p = p_o$

faţă de ipoteza alternativă $H_1 : p = p_1$

folosind testul secvenţial al raportului probabilităţilor cu α şi β daţi .

10. In urma cercetării numărului de ruperi ale firului, la un ring cu 224 fuse, s-au obţinut rezultatele din tabelul 7

T a b e l u l 7

Numărul de ruperi	0	1	2	3	4	5	6	7
Frecvenţa absolută	92	86	34	9	2	1	0	0

Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza că numărul de ruperi ale firului urmează repartiţia Poisson.

11. In urma cercetării unui lot de 130 şarje de fontă malleabilă cenuşie, din punctul de vedere al conţinutului de carbon, s-au obţinut rezultatele din tabelul 8

T a b e l u l 8

Intervale de grupare	1,70 - 1,90	1,90 - 2,00	2,00 - 2,10	2,10 - 2,20	2,20 - 2,30	2,30 - 2,40	2,40 - 2,50	2,50 - 2,60	2,60 - 2,70
Frecvenţa absolută	7	6	8	14	34	28	19	9	5

Să se testeze, la nivelul de semnificaţie $\alpha = 0,025$ ipoteza că conţinutul de carbon urmează repartiţia normală.

Scanned with OKEN Scanner

Capitolul VI

ANALIZA DISPERSIONALĂ

In unele probleme practice este necesar a şti dacă acţiunea unuia sau mai multor factori, asupra unei caracteristici este semnificativă sau întîmplătoare. Asemenea probleme se rezolvă cu ajutorul metodei analizei dispersionale introdusă de statisticianul R.A. Fisher.

Ideea fundamentală a analizei dispersionale constă în descompunerea unei dispersii de selecţie în mai multe componente independente, cu ajutorul cărora se construiesc criterii de verificare a ipotezelor cu privire la caracterul semnificativ sau întîmplător al acţiunii factorilor asupra caracteristicii cercetate. In cele ce urmează vom prezenta două modele de analiză dispersională în care acţiunea factorilor este deterministă.

§ 1. ANALIZA DISPERSIONALĂ UNIFACTORIALĂ.

Presupunem că asupra caracteristicii X urmărită la indivizii unei colectivităţi C, avînd repartiţia $N(m, \sigma^2)$, unde $m = M(X)$, $\sigma^2 = D^2(X)$ sînt necunoscuţi, acţionează factorul A cu variantele $A_1, \ldots, A_r$. Datorită acţiunii factorului A, colectivitatea C de caracteristică X, poate fi considerată ca descompusă în subcolectivităţile C_i de caracteristică X_i, avînd repartiţia $N(m_i, \sigma^2)$, $1 \leqslant i \leqslant r$. Caracteristicile X_i, $1 \leqslant i \leqslant r$, sînt independente. Din colectivitate se efectuează o selecţie repetată de volum n. Această selecţie poate fi considerată ca fiind constituită din r selecţii de volum, respecti

$n_1, n_2, \ldots, n_r$, efectuate respectiv din subcolectivităţile $C_1, C_2, \ldots, C_r$. Evident, $\sum_{i=1}^{r} n_i = n$. Pe baza selecţiei de volum n, trebuie verificată ipoteza

$$H : m_1 = m_2 = \ldots = m_r = m .$$

Acceptarea ipotezei H înseamnă că cele r selecţii provin dintr-o aceeaşi colectivitate de medie m şi dispersie σ^2.

Fie X_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant n_i$, variabilele aleatoare de selecţie, pe care le presupunem independente. Se consideră statisticile

$$(1) \qquad X_{i*} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} , \quad 1 \leqslant i \leqslant r , \qquad X_{**} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij}$$

Pentru i fixat, statistica X_{i*} reprezintă <u>media de selecţie a variantei A_i</u>. Statistica X_{**} reprezintă <u>media de selecţie totală</u>. Intrucît

$$(2) \qquad X_{**} = \frac{1}{n} \sum_{i=1}^{r} n_i X_{i*}$$

urmează că media de selecţie totală este egală cu media aritmetică ponderată a mediilor de selecţie corespunzătoare tuturor variantelor factorului A.

Vom demonstra că are loc

$$(3) \qquad \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{**})^2 = \sum_{i=1}^{r} n_i (X_{i*} - X_{**})^2 + \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{i*})^2 .$$

Intr-adevăr, avem

$$(4) \qquad \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{**})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left[(X_{i*} - X_{**}) + (X_{ij} - X_{i*}) \right]^2 =$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{i*} - X_{**})^2 + \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{i*})^2 + 2 \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{i*} - X_{**})(X_{ij} - X_{i*})$$

Vom arăta că are loc

$$(5) \qquad \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{i*} - X_{**})(X_{ij} - X_{i*}) = 0 .$$

Scanned with OKEN Scanner

Intr-adevăr, ţinînd seama de (1), avem

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{i*}-X_{**})(X_{ij}-X_{i*})=\sum_{i=1}^{r}(X_{i*}-X_{**})\sum_{j=1}^{n_i}(X_{ij}-X_{i*}) =$$

$$=\sum_{i=1}^{r}(X_{i*}-X_{**})(n_iX_{i*}-n_iX_{i*})=0 .$$

Avînd în vedere relaţia (5), relaţia (4) devine (3).

In urma notaţiilor

$$(6)\qquad Q(X,n)=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-X_{**})^2$$

$$Q_1(X,n)=\sum_{i=1}^{r}n_i(X_{i*}-X_{**})^2$$

$$Q_2(X,n)=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-X_{i*})^2$$

relaţia (3) devine

$$(7)\qquad Q(X,n)=Q_1(X,n)+Q_2(X,n) .$$

Teorema 1. Dacă ipoteza H se acceptă, atunci statisticile $Q_1(X,n)$ şi $Q_2(X,n)$ sînt independente, iar statisticile $Q(X,n)$, $Q_1(X,n)$ şi $Q_2(X,n)$ urmează repartiţia hi-pătrat, respectiv cu $n-1$, $r-1$, $n-r$ grade de libertate şi parametru σ .

Demonstraţie. Statistica $Q(X,n)$ se mai poate exprima prin

$$(8)\quad Q(X,n)=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-X_{**})^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left[(X_{ij}-m)-(X_{**}-m)\right]^2 =$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-m)^2+\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{**}-m)^2-2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-m)(X_{**}-m) =$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-m)^2+n(X_{**}-m)^2-2n(X_{**}-m)^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-m)^2-n(X_{**}-m)^2=$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-m)^2-\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}\frac{X_{ij}-m}{\sqrt{n}}\right]^2 .$$

Dacă notăm

$$X_{ij} - m = Y_{ij} , \quad 1 \leq i \leq r , \quad 1 \leq j \leq n_i$$

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} \frac{X_{ij} - m}{\sqrt{n}} = Z$$

atunci relaţia (8) se scrie sub forma

$$(9) \qquad Q(X,n) = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}^2 - Z^2 .$$

Intrucît variabilele aleatoare de selecţie X_{ij}, $1 \leq i \leq r$, $1 \leq j \leq n_i$, sînt independente şi urmează repartiţia $N(m, \sigma^2)$, rezultă că variabilele aleatoare Y_{ij}, $1 \leq i \leq r$, $1 \leq j \leq n_i$, sînt independente şi urmează repartiţia $N(0, \sigma^2)$. Variabila aleatoare Z reprezintă o formă liniară în variabilele aleatoare Y_{ij}, ai cărei coeficienţi satisfac relaţia

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(\frac{1}{\sqrt{n}}\right)^2 = 1 .$$

Conform teoremei 7, § 3, cap.I, variabila aleatoare Z urmează repartiţia $N(0, \sigma^2)$. Tinînd seama de (9) şi observaţia 1, §3, cap.I, rezultă că statistica Q(X,n) urmează o repartiţie hi-pătrat cu n-1 grade de libertate şi parametru σ .
Statistica

$$Q_1(X,n) = \sum_{i=1}^{r} n_i (X_{i*} - X_{**})^2$$

reprezintă suma pătratelor formelor liniare

$$L_i = \sqrt{n_i}(X_{i*} - X_{**})^2 , \quad 1 \leq i \leq r$$

ce satisfac relaţia

$$(10) \quad \sum_{i=1}^{r} \sqrt{n_i} L_i = \sum_{i=1}^{r} n_i (X_{i*} - X_{**}) = \sum_{i=1}^{r} n_i X_{i*} - nX_{**} = nX_{**} - nX_{**} = 0 .$$

Prin urmare, rangul lui $Q_1(X,n)$ este cel mult r-1. Statistica

$$Q_2(X,n) = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{i*})^2$$

reprezintă suma pătratelor formelor liniare

$$H_{ij} = X_{ij} - X_{i*},\ 1 \leq i \leq r,\ 1 \leq j \leq n_i$$

care satisfac relaţiile

$$(11)\quad \sum_{j=1}^{n_i} H_{ij} = \sum_{j=1}^{n_i} (X_{ij} - X_{i*}) = \sum_{j=1}^{n_i} X_{ij} - n_i X_{i*} = nX_{i*} - nX_{i*} = 0,\ 1 \leq i \leq r$$

Deci, rangul lui $Q_2(X,n)$ este cel mult n-r. Avînd în vedere (7) şi faptul că $Q(X,n)$ urmează o repartiţie hi-pătrat cu n-1 grade de libertate, rezultă că rangul lui $Q_1(X,n)$ este r-1, iar rangul lui $Q_2(X,n)$ este n-r. Conform teoremei 12, §3, cap.I, rezultă că $Q_1(X,n)$ şi $Q_2(X,n)$ sînt independente şi urmează respectiv repartiţiile hi-pătrat cu r-1 şi n-r grade de libertate, de parametru σ . Cu aceasta teorema este demonstrată.

Considerăm dispersiile de selecţie

$$(12)\quad S^2(X,n) = \frac{Q(X,n)}{n-1}$$ - dispersia de selecţie totală

$$S_1^2(X,n) = \frac{Q_1(X,n)}{r-1}$$ - dispersia de selecţie între variantele lui A

$$S_2^2(X,n) = \frac{Q_2(X,n)}{n-r}$$ - dispersia de selecţie reziduală .

Pe baza lui (12), relaţia (7) devine

$$(13)\quad (n-1)S^2(X,n) = (r-1)S_1^2(X,n) + (n-r)S_2^2(X,n)$$

Relaţia (13) constituie relaţia fundamentală a analizei dispersionale unifactoriale.

Teorema 2. Dacă ipoteza H se acceptă, atunci

$$M[S^2(X,n)] = M[S_1^2(X,n)] = M[S_2^2(X,n)] = \sigma^2 .$$

Demonstraţie. Vom face demonstraţia numai pentru $S^2(X,n)$ şi $S_1^2(X,n)$. Avem,

$$(14)\quad M[S^2(X,n)] = M\left[\frac{Q(X,n)}{n-1}\right] = \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} M(X_{ij} - X_{**})^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} M[(X_{ij} - m) - (X_{**} - m)]^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left[M(X_{ij}-m)^2 + M(X_{**}-m)^2 - 2M(X_{ij}-m)(X_{**}-m) \right]$$

Din faptul că ipoteza H se acceptă, rezultă

(15) $\quad M(X_{ij} - m)^2 = D^2(X_{ij}) = \sigma^2 \ , \ 1 \leq i \leq r \ , \ 1 \leq j \leq n_i \ .$

Conform teoremei 2, §3, cap.III și ținînd seama că ipoteza H se acceptă, se obține

(16) $\quad M(X_{**} - m)^2 = D^2(X_{**}) = \frac{D^2(X)}{n} = \frac{\sigma 2}{n} \ .$

Din independența variabilelor aleatoare de selecție $X_{ij} - m$, $1 \leq i \leq r$, $1 \leq j \leq n_i$, urmează

$$(17) \quad M\left[(X_{ij}-m)(X_{**}-m)\right] = M\left[(X_{ij}-m)\left(\frac{\sum_{k=1}^{r} \sum_{\ell=1}^{n_i} X_{k\ell}}{n} - m\right)\right] =$$

$$= \frac{1}{n} M\left[(X_{ij} - m)^2 + \sum_{(k,\ell)\neq(i,j)} \sum (X_{k\ell}-m)(X_{ij} - m)\right] =$$

$$= \frac{1}{n} D^2(X_{ij}) + \sum_{(k,\ell)\neq(i,j)} \sum M(X_{k\ell}-m)M(X_{ij}-m) = \frac{\sigma^2}{n}$$

deoarece $M(X_{k\ell} - m) = 0$, $1 \leq k \leq r$, $1 \leq \ell \leq n_i$.

Folosind (16) și (17), relația (14) devine

$$(18) \quad M\left[S^2(X,n)\right] = \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left[\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right] =$$

$$= \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \frac{n-1}{n} \sigma^2 = \sigma^2$$

Calculînd media lui $S_1^2(X,n)$, se obține

$$(19) \quad M\left[S_1^2(X,n)\right] = M\left[\frac{Q_1(X,n)}{r-1}\right] = \frac{1}{r-1} M\left[Q_1(X,n)\right] = \frac{1}{r-1} \sum_{i=1}^{r} n_i M(X_{i*}-X_{**})^2 =$$

$$= \frac{1}{r-1} \sum_{i=1}^{r} n_i M\left[(X_{i*} - m) - (X_{**} - m)\right]^2 =$$

$$= \frac{1}{r-1} \sum_{i=1}^{r} n_i M\left[(X_{i*}-m)^2 - 2(X_{i*}-m)(X_{**} - m) + (X_{**} - m)^2\right] =$$

$$= \frac{1}{r-1} \sum_{i=1}^{r} n_i \left[M(X_{i*}-m)^2 - 2M(X_{**}-m)(X_{i*}-m) + M(X_{**}-m)^2\right] =$$

$$= \frac{1}{r-1}\sum_{i=1}^{r} n_i\left[D^2(X_{i*})-2M(X_{**}-m)(X_{i*}-m) + D^2(X_{**})\right] .$$

Ținînd seama de teorema 2, §3, cap.III, avem

$$D^2(X_{i*}) = \frac{\sigma^2}{n_i} , \quad D^2(X_{**}) = \frac{\sigma^2}{n} . \tag{20}$$

Calculînd $M(X_{**} - m)(X_{i*} - m)$, se obține

$$M(X_{**}-m)(X_{i*}-m) = M\left[\left(\sum_{j=1}^{r} \frac{n_j X_{j*}}{n} - m\right)(X_{i*} - m)\right] = \tag{21}$$

$$= M\left[\left(\sum_{j=1}^{r} n_j \frac{X_{j*} - m}{n}\right)(X_{i*} - m)\right] =$$

$$= \frac{1}{n} M\left[n_i(X_{i*}-m)^2 + \sum_{j=1}^{r} n_j(X_{j*} - m)(X_{i*} - m)\right] =$$

$$= \frac{1}{n} n_i M(X_{i*}-m)^2 + \frac{1}{n}\sum_{j=1}^{r} n_j M\left[(X_{j*} - m)(X_{i*} - m)\right] =$$

$$= \frac{n_i}{n} D^2(X_{i*}) + \frac{1}{n}\sum_{j=1}^{r} n_j M(X_{j*} - m)M(X_{i*} - m) .$$

Ținînd seama de (20) și de independența variabilelor aleatoare $X_{i*} - m$, $1 \leq i \leq r$, relația (21) devine

$$M\left[(X_{**} - m)(X_{i*} - m)\right] = \frac{n_i \sigma^2}{n n_i} = \frac{\sigma^2}{n} . \tag{22}$$

Folosind (20) și (21), relația (19) devine

$$M\left[S_1^2(X,n)\right] = \frac{1}{r-1}\sum_{i=1}^{r} n_i\left[\frac{\sigma^2}{n_i} - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n}\right] = \tag{23}$$

$$= \frac{1}{r-1}\sum_{i=1}^{r}\left(\sigma^2 - \frac{n_i\sigma^2}{n}\right) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2 .$$

Analog se calculează $M\left[S_2^2(X,n)\right]$, fără presupunerea că ipoteza H se acceptă. Cu aceasta teorema este demonstrată.

Considerăm statistica

$$F(X,n) = \frac{n-r}{r-1} \cdot \frac{Q_1(X,n)}{Q_2(X,n)} = \frac{S_1^2(X,n)}{S_2^2(X,n)} .$$

Scanned with OKEN Scanner

Conform teoremei 14, §3, cap.I, statistica $F(X,n)$ urmează repartiţia Snedecor cu $r-1$ şi $n-r$ grade de libertate.

Fie valorile de selecţie x_{ij} , $1 \leq i \leq r$, $1 \leq j \leq n_i$. Notăm

$$x_{i*} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad 1 \leq i \leq r , \quad x_{**} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}$$

$$\bar{Q} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij}-x_{**})^2, \quad \bar{Q}_1 = \sum_{i=1}^{r} n_i (x_{i*}-x_{**})^2, \quad \bar{Q}_2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij}-x_{i*})^2$$

$$s_1^2 = \frac{\bar{Q}_1}{r-1} , \quad s_2^2 = \frac{\bar{Q}_2}{n-r} , \quad s^2 = \frac{\bar{Q}}{n-1}$$

Valoarea statisticii $F(X,n)$ pentru selecţia efectuată este

$$F_o = \frac{s_1^2}{s_2^2} .$$

După alegerea nivelului de semnificaţie α , cu ajutorul tabelelor VIII sau IX din anexă, se determină valoarea critică F_α, astfel încît să aibă loc

$$P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F(X,n) > F_\alpha\right) = \alpha .$$

Regiunea critică a testului este

$$U = \left\{ x_{ij} , 1 \leq i \leq r , 1 \leq j \leq n_i : \frac{s_1^2}{s_2^2} > F_\alpha \right\}$$

iar mulţimea din R, corespunzătoare lui U este $I = (F_\alpha, +\infty)$

Testul este următorul :

i) dacă $F_o \leq F_\alpha$, ipoteza H se acceptă

ii) dacă $F_o > F_\alpha$, ipoteza H se respinge .

Observaţia 1. In calculul lui F_o se consideră la numărător dispersia cea mai mare dintre s_1^2 şi s_2^2 .

Observaţia 2. Media colectivităţii m poate fi estimată de X_{**} , iar dispersia colectivităţii σ^2 poate fi estimată de oricare din dispersiile de selecţie $S^2(X,n)$, $S_1^2(X,n)$, $S_2^2(X,n)$.

Observaţia 3. Analiza dispersională unifactorială cu două variante revine la un test t prezentat în §8, cap.V.

Observația 4. Ținînd seama de identitatea

$$\frac{1}{N}\sum_{i=1}^{N}(Z_i - Z_*)^2 = \frac{1}{N}\sum_{i=1}^{N}Z_i^2 - \frac{(\sum_{i=1}^{N}Z_i)^2}{N^2}$$

sumele de pătrate $\bar{Q}, \bar{Q}_1, \bar{Q}_2$ se pot exprima prin

$$\bar{Q} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij}^2 - \frac{(\sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij})^2}{n}$$

$$\bar{Q}_1 = \sum_{i=1}^{r}\frac{(\sum_{j=1}^{n_i}x_{ij})^2}{n_i} - \frac{(\sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij})^2}{n}$$

$$\bar{Q}_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij}^2 - \sum_{i=1}^{r}\frac{(\sum_{j=1}^{n_i}x_{ij})^2}{n_i}\ .$$

In aplicații, se folosește următoarea schemă de analiză dispersională unifactorială :

Tabelul 9

Dispersia statistică	Sume de pătrate	Numărul gradelor de libertate	Estimații pentru σ^2	Valoarea statisticii
între variantele factorului A	$\bar{Q}_1$	r-1	$s_1^2 = \frac{\bar{Q}_1}{r-1}$	$F_o = \frac{s_1^2}{s_2^2}$
reziduală	$\bar{Q}_2$	n-r	$s_2^2 = \frac{\bar{Q}_2}{n-r}$	
totală	$\bar{Q}$	n-1	$s^2 = \frac{\bar{Q}}{n-1}$	

Scanned with OKEN Scanner

§ 2. ANALIZA DISPERSIONALĂ BIFACTORIALĂ FĂRĂ INTERACŢIUNE ÎNTRE FACTORI.

Fie X caracteristica urmărită la indivizii unei colectivităţi avînd repartiţia $N(m, \sigma^2)$, unde $m = M(X)$, $\sigma^2 = D^2(X)$ sînt necunoscuţi. Presupunem că asupra lui X acţionează simultan, dar independent, factorul A cu variantele $A_1, \ldots, A_i, \ldots, A_r$ şi factorul B cu variantele $B_1, \ldots, B_j, \ldots, B_v$. Datorită acţiunii celor doi factori, colectivitatea poate fi considerată ca descompusă în $r\,v$ subcolectivităţi avînd repartiţiile $N(m_{ij}, \sigma^2)$, $1 \leq i \leq r$, $1 \leq j \leq v$. Din colectivitate se efectuează o selecţie de volum $r\,v$, selecţie ce poate fi considerată ca formată din $r\,v$ selecţii, de volum 1, cîte una din fiecare subcolectivitate. Aceasta înseamnă că, pentru fiecare combinaţie de variante A_iB_j ale celor doi factori, se efectuează o singură observaţie. Fie X_{ij} variabilele aleatoare de selecţie, independente, urmînd repartiţiile $N(m_{ij}, \sigma^2)$, $1 \leq i \leq r$, $1 \leq j \leq v$. Pe baza selecţiei efectuate, de volum $r\,v$, trebuie verificată ipoteza

$$H : m_{ij} = m \quad , \quad 1 \leq i \leq r \quad , \quad 1 \leq j \leq v \; .$$

Acceptarea ipotezei H înseamnă că cele $r\,v$ selecţii, de volum 1, provin dintr-o aceeaşi colectivitate de medie m şi dispersie σ^2.

Se consideră statisticile

$$(1) \qquad \begin{cases} X_{i*} = \frac{1}{v} \sum_{j=1}^{v} X_{ij} \; , & 1 \leq i \leq r \\ X_{*j} = \frac{1}{r} \sum_{i=1}^{r} X_{ij} \; , & 1 \leq j \leq v \\ X_{**} = \frac{1}{rv} \sum_{i=1}^{r} \sum_{j=1}^{v} X_{ij} \; . \end{cases}$$

Pentru i fixat, statistica X_{i*} reprezintă media de selecţie a variantei A_i a factorului A . Pentru j fixat, statistica X_{*j} reprezintă media de selecţie a variantei B_j a factorului B. Statistica X_{**} reprezintă media de selecţie totală.

Intrucît

$$X_{**} = \frac{1}{r}\sum_{i=1}^{r} X_{i*} = \frac{1}{v}\sum_{j=1}^{v} X_{*j} \tag{2}$$

urmează că, media de selecţie totală este egală cu media aritmetică a mediilor corespunzătoare tuturor variantelor lui A sau variantelor lui B .

Vom demonstra

$$\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{**})^2 = v\sum_{i=1}^{r}(X_{i*}-X_{**})^2 + r\sum_{j=1}^{v}(X_{*j}-X_{**})^2 + \tag{3}$$

$$+ \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij} - X_{i*} - X_{*j} + X_{**})^2 .$$

Intr-adevăr, avem

$$\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{**})^2 = \sum_{i=1}^{r}\sum_{j=1}^{v}\left[(X_{i*}-X_{**})+(X_{*j}-X_{**})+(X_{ij}-X_{i*}-X_{*j}+X_{**})\right]^2 = \tag{4}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*}-X_{**})^2 + \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{*j}-X_{**})^2 + \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{i*}-X_{*j}+X_{**})^2 +$$

$$+2\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*}-X_{**})(X_{*j}-X_{**}) + 2\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*}-X_{**})(X_{ij}-X_{i*}-X_{*j}+X_{**}) +$$

$$+ 2\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{*j} - X_{**})(X_{ij} - X_{i*} - X_{*j} + X_{**}) .$$

Vom arăta că au loc relaţiile :

$$\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*} - X_{**})(X_{*j} - X_{**}) = 0 \tag{5}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*} - X_{**})(X_{ij} - X_{i*} - X_{*j} + X_{**}) = 0 \tag{6}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{*j} - X_{**})(X_{ij} - X_{i*} - X_{*j} + X_{**}) = 0 . \tag{7}$$

Scanned with OKEN Scanner

Intr-adevăr, ţinînd seama de (2), se obţine

$$(8)\quad \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*}-X_{**})(X_{*j}-X_{**}) = \sum_{i=1}^{r}(X_{i*}-X_{**})\sum_{j=1}^{v}(X_{*j}-X_{**}) =$$

$$= (rX_{**}-rX_{**})(vX_{**}-vX_{**}) = 0 .$$

Ţinînd seama de (1), se obţine

$$(9)\quad \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{i*}-X_{**})(X_{ij}-X_{i*}-X_{*j}+X_{**}) =$$

$$= \sum_{i=1}^{r}(X_{i*}-X_{*})\sum_{j=1}^{v}(X_{ij}-X_{i*}-X_{*j}+X_{**}) =$$

$$= \sum_{i=1}^{r}(X_{i*}-X_{**})(vX_{i*}-vX_{i*}-vX_{**}+vX_{**}) = 0 .$$

Analog cu relaţia (6) se demonstrează relaţia (7). Avîmd în vedere relaţiile (5), (6), (7), relaţia (4) se reduce la relaţia (3).

In urma notaţiilor

$$Q(X,n) = \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{**})^2,\quad Q_1(X,n) = v\sum_{i=1}^{r}(X_{i*}-X_{**})^2$$

$$Q_2(X,n) = r\sum_{j=1}^{v}(X_{*j}-X_{**})^2,\quad Q_3(X,n) = \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{i*}-X_{*j}+X_{**})^2$$

relaţia (3), devine

$$(10)\qquad Q(X,n) = Q_1(X,n) + Q_2(X,n) + Q_3(X,n)$$

<u>Teorema 1</u>. Dacă ipoteza H se acceptă, atunci statisticile $Q_1(X,n)$, $Q_2(X,n)$ şi $Q_3(X,n)$ sînt independente, iar statisticile $Q(X,n)$, $Q_1(X,n)$, $Q_2(X,n)$, $Q_3(X,n)$ urmează repartiţia hi-pătrat, respectiv cu $rv-1, r-1, v-1, (r-1)(v-1)$ grade de libertate şi parametru σ .

<u>Demonstraţie</u>. Statistica $Q(X,n)$ se mai poate exprima prin

$$(11)\quad Q(X,n) = \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{**})^2 = \sum_{i=1}^{r}\sum_{j=1}^{v}\left[(X_{ij}-m)-(X_{**}-m)\right]^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)^2 + \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{**}-m)^2 - 2\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)(X_{**}-m) =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)^2 + rv(X_{**}-m)^2 - 2rv(X_{**}-m)^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)^2 - rv(X_{**}-m)^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)^2 - rv\left(\frac{\sum_{i=1}^{r}\sum_{j=1}^{v}X_{ij}}{r\,v} - m\right)^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)^2 - \left[\frac{\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-m)}{\sqrt{r\,v}}\right]^2$$

Dacă notăm

$$X_{ij} - m = Y_{ij}\ ,\ 1 \leqslant i \leqslant r\ ,\ 1 \leqslant j \leqslant v$$

$$\sum_{i=1}^{r}\sum_{j=1}^{v}\frac{X_{ij}-m}{\sqrt{r\,v}} = Z$$

relaţia (11) se scrie sub forma

$$(12) \qquad Q(X,n) = \sum_{i=1}^{r}\sum_{j=1}^{v}Y_{ij}^2 - Z^2 \ .$$

Intrucît variabilele aleatoare de selecţie X_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant v$, sînt independente şi urmează repartiţia $N(m,\sigma^2)$, rezultă că variabilele aleatoare Y_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant v$, sînt independente şi urmează repartiţia $N(0,\sigma^2)$. Variabila aleatoare Z reprezintă o formă liniară în variabilele aleatoare Y_{ij}, ai cărei coeficienţi satisfac relaţia

$$\sum_{i=1}^{r}\sum_{j=1}^{v}\left(\frac{1}{\sqrt{r\,v}}\right)^2 = 1 \ .$$

Conform teoremei 7, §3, cap.I, variabila aleatoare Z urmează repartiţia $N(0,\sigma^2)$. Tinînd seama de (12) şi observaţia 1, §3,

Scanned with OKEN Scanner

cap.I, rezultă că statistica $Q(X,n)$ urmează o repartiție bi-pătrat cu $rv-1$ grade de libertate și parametru σ . Statistica

$$Q_1(X,n) = v \sum_{i=1}^{r} (X_{i*} - X_{**})^2$$

reprezintă suma pătratelor formelor liniare

$$L_i = \sqrt{v}\ (X_{i*} - X_{**}) \ , \ 1 \leq i \leq r$$

ce satisfac relația

$$(13) \quad \sum_{i=1}^{r} \sqrt{v}\ L_i = \sum_{i=1}^{r} v(X_{i*} - X_{**}) = vrX_{**} - vrX_{**} = 0 \ .$$

Prin urmare, rangul lui $Q_1(X,n)$ este cel mult r-1. Statistica

$$Q_2(X,n) = r \sum_{j=1}^{v} (X_{*j} - X_{**})^2$$

reprezintă suma pătratelor formelor liniare

$$K_j = \sqrt{r}(X_{*j} - X_{**}) \ , \quad 1 \leq j \leq v$$

ce satisfac relația

$$(14) \quad \sum_{j=1}^{v} \sqrt{r}\ K_j = \sum_{j=1}^{v} r(X_{*j} - X_{**}) = rvX_{**} - rvX_{**} = 0$$

Prin urmare, rangul lui $Q_2(X,n)$ este cel mult v-1. Statistica

$$Q_3(X,n) = \sum_{i=1}^{r} \sum_{j=1}^{v} (X_{ij} - X_{i*} - X_{*j} + X_{**})^2$$

reprezintă suma pătratelor formelor liniare

$$H_{ij} = X_{ij} - X_{i*} - X_{*j} + X_{**} \ , \ 1 \leq i \leq r \ , \ 1 \leq j \leq v$$

ce satisfac relațiile

$$(15) \quad \sum_{i=1}^{r} H_{ij} = \sum_{i=1}^{r} (X_{ij} - X_{i*} - X_{*j} + X_{**}) = rX_{*j} - rX_{**} - rX_{*j} + rX_{**} = 0, \ 1 \leq j \leq v$$

$$(16) \quad \sum_{j=1}^{v} H_{ij} = \sum_{j=1}^{v} (X_{ij} - X_{i*} - X_{*j} + X_{**}) = vX_{i*} - vX_{i*} - vX_{**} + vX_{**} = 0, \ 1 \leq i \leq r$$

Deoarece pentru $i = j$, relațiile

$$\sum_{i=1}^{r} (X_{ij} - X_{i*} - X_{*j} + X_{**}) = 0, \quad \sum_{j=1}^{v} (X_{ij} - X_{i*} - X_{*j} + X_{**}) = 0$$

coincid, urmează că rangul lui $Q_3(X,n)$ este cel mult $rv-(r+v-1)=(r-1)(v-1)$. Avînd în vedere (10) şi faptul că $Q(X,n)$ urmează o repartiţie hi-pătrat cu $rv-1$ grade de libertate, rezultă că $Q_1(X,n)$, $Q_2(X,n)$, $Q_3(X,n)$ au rangul respectiv, $r-1, v-1, (r-1)(v-1)$.

Conform teoremei 12, §3, cap.I, se obţine că $Q_1(X,n)$, $Q_2(X,n)$, $Q_3(X,n)$ sînt independente şi urmează repartiţia hi-pătrat, respectiv cu $r-1$, $v-1$, $(r-1)(v-1)$ grade de libertate şi parametru σ . Cu aceasta teorema este demonstrată.

Dacă se consideră dispersiile de selecţie

$$S^2(X,n) = \frac{Q(X,n)}{rv-1}$$ - dispersia de selecţie totală

$$S_1^2(X,n) = \frac{Q_1(X,n)}{r-1}$$ - dispersia de selecţie între variantele factorului A

$$S_2^2(X,n) = \frac{Q_2(X,n)}{v-1}$$ - dispersia de selecţie între variantele factorului B

$$S_3^2(X,n) = \frac{Q_3(X,n)}{(r-1)(v-1)}$$ - dispersia de selecţie reziduală

atunci relaţia (10) se scrie sub forma

$$(17)\quad (rv-1)S^2(X,n)=(r-1)S_1^2(X,n)+(v-1)S_2^2(X,n)+(r-1)(v-1)S_3^2(X,n)$$

Relaţia (17) constituie relaţia fundamentală a analizei dispersionale bifactoriale, fără interacţiune între factori .

Teorema 2. Dacă ipoteza H se acceptă, atunci

$$M[S^2(X,n)]= M[S_1^2(X,n)]= M[S_2^2(X,n)]= M[S_3^2(X,n)] = \sigma^2 .$$

Demonstraţie. Vom face demonstraţia numai pentru $S^2(X,n)$, $S_1^2(X,n)$ şi $S_2^2(X,n)$. Dacă ipoteza H se acceptă, atunci

$M(X_{ij}) = m$, $1 \leq i \leq r$, $1 \leq j \leq v$, $M(X_{i*}) = m$, $1 \leq i \leq r$

$M(X_{*j}) = m$, $1 \leq j \leq v$, $M(X_{**}) = m$.

Calculînd media lui $S^2(X,n)$, se obţine

$$(18)\qquad M[S^2(X,n)]= M\left[\frac{Q(X,n)}{rv-1}\right] = \frac{1}{rv-1}\left[M\, Q(X,n)\right] =$$

$$= \frac{1}{rv-1} M\Big[\sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij}-X_{**})^2\Big] = \frac{1}{rv-1}\sum_{i=1}^{r}\sum_{j=1}^{v} M(X_{ij}-X_{**})^2$$

Dacă ipoteza H se acceptă, atunci

$$M(X_{ij} - X_{**}) = M(X_{ij}) - M(X_{**}) = 0, \quad 1 \leq i \leq r, \quad 1 \leq j \leq v$$

Calculăm

(19) $$M(X_{ij} - X_{**})^2 = M\big[(X_{ij} - m) - (X_{**} - m)\big]^2 =$$

$$= M\big[(X_{ij}-m)^2 - 2(X_{ij}-m)(X_{**}-m) + (X_{**} - m)^2\big] =$$

$$= M\Big[(X_{ij}-m)^2 - 2(X_{ij}-m)\Big(\frac{\sum_{k=1}^{r}\sum_{\ell=1}^{v} X_{k\ell}}{rv} - m\Big) + (X_{**} - m)^2\Big] =$$

$$= M(X_{ij}-m)^2 - \frac{2}{rv}M(X_{ij}-m)^2 - \frac{2}{rv}\sum_{(i,j)\neq(k,\ell)} M(X_{ij}-m)(X_{k\ell}-m) +$$

$$+ M(X_{**} - m)^2 = D^2(X_{ij}) - \frac{2}{rv}D^2(X_{ij}) + D^2(X_{**})$$

întrucît $M(X_{ij} - m) = 0$, $1 \leq i \leq r$, $1 \leq j \leq v$ şi variabilele aleatoare $X_{ij} - m$, $1 \leq i \leq r$, $1 \leq j \leq v$, sînt independente. Ţinînd seama de teorema 2, §3, cap.III, rezultă

$$D^2(X_{**}) = \frac{\sigma^2}{rv}$$

Prin urmare, relaţia (19) devine

(20) $$M\big[S^2(X,n)\big] = \frac{1}{rv-1}\, rv\, \frac{(rv-1)\sigma^2}{rv} = \sigma^2$$

Calculăm

(21) $$M\big[S_1^2(X,n)\big] = M\Big[\frac{v}{r-1}\sum_{i=1}^{r}(X_{i*}-X_{**})^2\Big] = \frac{v}{r-1}\sum_{i=1}^{r} M(X_{i*}-X_{**})^2$$

Dacă ipoteza H se acceptă, atunci

(22) $$M(X_{i*} - X_{**})^2 = M\big[(X_{i*}-m) - (X_{**} - m)\big]^2 =$$

$$= M\big[(X_{i*}-m)^2 - 2(X_{i*}-m)(X_{**}-m) + (X_{**} - m)^2\big] =$$

$$= M\Big[(X_{i*}-m)^2 - 2(X_{i*}-m)\Big(\frac{\sum_{k=1}^{r} X_{k*}}{r} - m\Big) + (X_{**} - m)^2\Big] =$$

$$= M(X_{i*}-m)^2 - \frac{2}{r}M(X_{i*}-m)^2 - \frac{2}{r}\sum_{i\neq k} M(X_{i*}-m)(X_{k*}-m) + M(X_{**}-m)^2 =$$

$$= D^2(X_{i*}) - \frac{2}{r}D^2(X_{i*}) + D^2(X_{**}) .$$

Din teorema 2, §3, cap.III, rezultă

$$D^2(X_{i*}) = \frac{\sigma^2}{v} , \quad 1 \leq i \leq r . \tag{23}$$

Folosind (23) și faptul că $D^2(X_{**}) = \frac{\sigma^2}{rv}$, relația (22) devine

$$M(X_{i*} - X_{**})^2 = \frac{2}{v} - \frac{2}{r}\frac{\sigma^2}{v} + \frac{\sigma^2}{rv} = \frac{(r-1)\sigma^2}{rv} . \tag{24}$$

Avînd în vedere relația (24), relația (21) devine

$$M\left[S_1^2(X,n)\right] = \frac{v}{r-1} \cdot r \cdot \frac{(r-1)\sigma^2}{rv} = \sigma^2 \tag{25}$$

Calculăm

$$M\left[S_2^2(X,n)\right] = \frac{r}{v-1}\sum_{j=1}^{v} M(X_{*j}-X_{**})^2 = \frac{r}{v-1}\sum_{j=1}^{v} M\left[(X_{*j}-m)-(X_{**}-m)\right]^2 = \tag{26}$$

$$= \frac{r}{v-1}\sum_{j=1}^{v} M\left[(X_{*j}-m)^2 - 2(X_{*j}-m)(X_{**}-m) + (X_{**}-m)^2\right] =$$

$$= \frac{r}{v-1}\sum_{j=1}^{v}\left[M(X_{*j}-m)^2 - 2M(X_{*j}-m)\left(\frac{\sum_{\ell=1}^{v} X_{*\ell}}{v} - m\right) + M(X_{**}-m)^2\right] =$$

$$= \frac{r}{v-1}\sum_{j=1}^{v}\left[M(X_{*j}-m)^2 - \frac{2}{v}M(X_{*j}-m)^2 - \frac{2}{v}\sum_{j\neq\ell} M(X_{*\ell}-m)(X_{*j}-m) + M(X_{**}-m)^2\right] =$$

$$= \frac{r}{v-1}\sum_{j=1}^{v}\left[D^2(X_{*j}) - \frac{2}{v}D^2(X_{*j}) + D^2(X_{**})\right] .$$

Pe baza teoremei 2, §3, cap.III, avem

$$D^2(X_{*j}) = \frac{\sigma^2}{r} , \quad 1 \leq j \leq v . \tag{27}$$

Ținînd seama de (27) și faptul că $D^2(X_{**}) = \frac{\sigma^2}{rv}$, relația (26) devine

$$M\left[S_2^2(X,n)\right] = \frac{r}{v-1}\sum_{j=1}^{v}\left[\frac{\sigma^2}{r} - \frac{2}{v}\frac{\sigma^2}{r} + \frac{\sigma^2}{rv}\right] = \sigma^2 . \tag{28}$$

Folosind relația

$$S_3^2(X,n) = \frac{1}{(r-1)(v-1)}\sum_{i=1}^{r}\sum_{j=1}^{v}\left[(X_{ij}-X_{**})-(X_{i*}-X_{**})-(X_{*j}-X_{**})\right]^2 \tag{29}$$

se poate demonstra că

(30) $$M\left[s_3^2(X,n)\right] = \sigma^2 .$$

Cu aceasta teorema este demonstrată.

Considerăm statisticile

$$F_A(X,n) = \frac{\frac{Q_1(X,n)}{r-1}}{\frac{Q_3(X,n)}{(r-1)(v-1)}} = \frac{s_1^2(X,n)}{s_3^2(X,n)} = (v-1)\frac{Q_1(X,n)}{Q_3(X,n)}$$

$$F_B(X,n) = \frac{\frac{Q_2(X,n)}{v-1}}{\frac{Q_3(X,n)}{(r-1)(v-1)}} = \frac{s_2^2(X,n)}{s_3^2(X,n)} = (r-1)\frac{Q_2(X,n)}{Q_3(X,n)} .$$

Conform teoremei 14, §3, cap.I, statistica $F_A(X,n)$ urmează repartiţia Snedecor cu r-1 şi (r-1)(v-1) grade de libertate, iar statistica $F_B(X,n)$ urmează repartiţia Snedecor cu v-1 şi (r-1)(v-1) grade de libertate de libertate. Fie valorile de selecţie x_{ij} , $1 \leq i \leq r$, $1 \leq j \leq v$.

Notăm

$$x_{i*} = \frac{1}{v}\sum_{j=1}^{v} x_{ij},\ 1 \leq i \leq r,\quad x_{*j} = \frac{1}{r}\sum_{i=1}^{r} x_{ij},\ 1 \leq j \leq v,\quad x_{**} = \frac{1}{rv}\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}$$

$$\bar{Q} = \sum_{i=1}^{r}\sum_{j=1}^{v}(x_{ij}-x_{**})^2,\ \bar{Q}_1 = v\sum_{i=1}^{r}(x_{i*}-x_{**})^2,\ \bar{Q}_2 = r\sum_{j=1}^{v}(x_{*j}-x_{**})^2$$

$$\bar{Q}_3 = \sum_{i=1}^{r}\sum_{j=1}^{v}(x_{ij}-x_{i*}-x_{*j}+x_{**})^2,\ s^2 = \frac{\bar{Q}}{rv-1},\ s_1^2 = \frac{\bar{Q}_1}{r-1},\ s_2^2 = \frac{\bar{Q}_2}{v-1}$$

$$s_3^2 = \frac{\bar{Q}_3}{(r-1)(v-1)}$$

Valorile statisticilor $F_A(X,n)$, $F_B(X,n)$ pentru selecţia considerată sînt respectiv $F_A^o = \frac{s_1^2}{s_3^2}$, $F_B^o = \frac{s_2^2}{s_3^2}$.

Notăm

$$m_{i*} = \frac{1}{v}\sum_{j=1}^{v} m_{ij} \quad , \quad m_{*j} = \frac{1}{r}\sum_{i=1}^{r} m_{ij}$$

şi considerăm ipotezele

$$H_A : m_{1*} = m_{2*} = \ldots = m_{r*} \qquad H_B : m_{*1} = m_{*2} = \ldots = m_{*v}$$

Evident, dacă ipoteza H se acceptă, atunci şi ipotezele H_A şi H_B se acceptă.

Pentru nivelul de semnificaţie α fixat, cu ajutorul tabelului VIII sau IX din anexă, se determină valoarea critică $F_{A;\alpha}$, astfel încît să aibă loc

$$P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : F_A(X,n) > F_{A;\alpha}\right) = \alpha \quad .$$

Regiunea critică a testului pentru verificarea ipotezei H_A este

$$U = \left\{ x_{ij} \ , \ 1 \leq i \leq r \ , \ 1 \leq j \leq v : F_A^o > F_{A;\alpha} \right\}$$

iar mulţimea din R , corespunzătoare lui U este

$$I = (F_{A;\alpha}, +\infty) \ .$$

Testul pentru verificarea ipotezei H_A este :

i) dacă $F_A^o \leq F_{A;\alpha}$, ipoteza H_A se acceptă

ii) dacă $F_A^o > F_{A;\alpha}$, ipoteza H_A se respinge .

Acceptarea ipotezei H_A, înseamnă că variantele factorului A influenţează întîmplător caracteristica X. Analog se face testarea ipotezei H_B.

Observaţia 1. Media colectivităţii m poate fi estimată de X_{**}, iar dispersia colectivităţii σ^2 poate fi estimată de oricare dintre dispersiile de selecţie $S^2(X,n), S_1^2(X,n)$, $S_2^2(X,n)$, $S_3^2(X,n)$.

Observaţia 2. Ţinînd seama de identitatea

$$\frac{1}{N}\sum_{i=1}^{N}(Z_i - Z_*)^2 = \frac{1}{N}\sum_{i=1}^{N} Z_i^2 - \frac{(\sum_{i=1}^{N} Z_i)^2}{N^2}$$

sumele de pătrate $\bar{Q}$, $\bar{Q}_1$, $\bar{Q}_2$, $\bar{Q}_3$ se exprimă prin

$$\bar{Q} = \sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}^2 - \frac{1}{rv}\left(\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}\right)^2$$

$$\bar{Q}_1 = \frac{1}{v}\sum_{i=1}^{r}\left(\sum_{j=1}^{v} x_{ij}\right)^2 - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}\right)^2}{rv}$$

$$\bar{Q}_2 = \frac{1}{r}\sum_{j=1}^{v}\left(\sum_{i=1}^{r} x_{ij}\right)^2 - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}\right)^2}{rv}$$

$$\bar{Q}_3 = \sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}^2 - \frac{1}{v}\sum_{i=1}^{r}\left(\sum_{j=1}^{v} x_{ij}\right)^2 - \frac{1}{r}\sum_{j=1}^{v}\left(\sum_{i=1}^{r} x_{ij}\right)^2 + \frac{1}{rv}\left(\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}\right)^2$$

In aplicaţii, se foloseşte următoarea schemă de analiză dispersională bifactorială fără interacţiune între factori :

T a b e l u l 10

Dispersia statistică	Sume de pătrate	Numărul gradelor de libertate	Estimaţii pentru σ^2	Valoarea statisticii
între variantele factorului A	$\bar{Q}_1$	r-1	$s_1^2 = \frac{\bar{Q}_1}{r-1}$	$F_A^o = \frac{s_1^2}{s_3^2}$
între variantele factorului B	$\bar{Q}_2$	v-1	$s_2^2 = \frac{\bar{Q}_2}{v-1}$	$F_B^o = \frac{s_2^2}{s_3^2}$
reziduală	$\bar{Q}_3$	(r-1)(v-1)	$s_3^2 = \frac{\bar{Q}_3}{(r-1)(v-1)}$	
totală	$\bar{Q}$	rv-1	$s^2 = \frac{Q}{rv-1}$	

P R O B L E M E

1. O fabrică produce lămpi electrice. Pe baza unor experienţe s-a constatat că durata de ardere a lămpilor electrice este o caracteristică ce urmează o repartiţie normală. Fabrica dispune de trei procedee tehnologice de fabricare a lămpilor electrice. Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza că, durata medie de ardere a lămpilor electrice este aceeaşi pentru cele trei procedee tehnologice, folosind datele din tabelul 11

T a b e l u l 11

Procedeu tehnologic	Durata de ardere						
I	1600	1610	1630	1680	1700	1710	1800
II	1580	1640	1650	1700	1750		
III	1510	1520	1530	1570	1600	1680	

2. Pentru determinarea numărului de bacterii din sol, un l de pămînt, a fost împărţit în patru parcele egale şi în fiecar parcelă au fost făcute însămînţări de bacterii pe plăci. Se co sideră că numărul de bacterii dezvoltate pe o parcelă este o c racteristică ce urmează repartiţia normală. Să se testeze, l nivelul de semnificaţie $\alpha = 0,05$, ipoteza că, numărul mediu d bacterii din cele patru parcele este acelaşi, folosind datele din tabelul 12.

T a b e l u l 12

Placa / Parcela	1	2	3	4	5	6	7
I	72	69	63	59	53	59	51
II	74	69	70	69	58	66	52
III	78	74	70	58	56	58	56
IV	69	67	66	62	58	62	54

3. Randamentul unei reacţii de conversie este o caracteristică ce urmează repartiţia normală. In urma experimentării reacţiei de conversie în sistem continuu la două nivele de temperatură şi două debite de alimentare, s-au obţinut rezultatele din tabelul 13

T a b e l u l 13

Debit / Temperatura	D_1	D_2
T_1	43,21	45,16
T_2	44,05	44,3

Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipotezele cu privire la semnificaţia influenţei temperaturii şi debitului asupra randamentului reacţiei de conversie.

4. Timpul de polimerizare este o caracteristică ce urmează repartiţia normală. Intr-o reacţie de polimerizare se încearcă patru catalizatori la trei temperaturi. Să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipotezele cu privire la semnificaţia catalizatorului şi temperaturii asupra timpului de polimerizare, folosind datele din tabelul 14

T a b e l u l 14

Catalizator / Temperatura	I	II	III	IV
25	25	28	22	24
50	27	29	23	23
75	30	32	26	29

Scanned with OKEN Scanner

C a p i t o l u l VII

CORELAȚIA ȘI REGRESIA SIMPLĂ

Scanned with OKEN Scanner

§ 1. REPARTIŢII BROBABILISTICE BIDIMENSIONALE

Fie $V = (X,Y)$ un vector aleator bidimensional definit pe cîmpul de probabilitate $\{\Omega, \mathcal{K}, P\}$. Dacă V urmează o repartiţie probabilistică de tip discret, atunci aceasta este complet determinată de tabloul de repartiţie

$$\left[\begin{matrix}(x_i,y_j)\\ p_{ij}\end{matrix}\right]_{\substack{i\in I\\ j\in J}}$$

unde $p_{ij} = P\left[\{\omega\in\Omega : X(\omega) = x_i\}\cap\{\omega\in\Omega : Y(\omega) = y_j\}\right]$, I şi J sînt mulţimi cel mult numărabile şi $\sum_{i\in I}\sum_{j\in J} p_{ij} = 1$.

In cele ce urmează vom presupune că $I = \{1,2,\dots,r\}$, $J = \{1,2,\dots,s\}$. Dacă V urmează o repartiţie probabilistică de tip continuu, atunci aceasta este complet determinată de densitatea de repartiţie $f(x,y)$ care satisface condiţiile

$$f(x,y) \geqslant 0$$

oricare ar fi $(x,y)\in R^2$ şi

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1 .$$

Dacă V urmează o repartiţie probabilistică de tip discret, atunci probabilităţile marginale se definesc prin

$$(1)\qquad p_{i*} = \sum_{j=1}^{s} p_{ij} , \quad 1\leqslant i\leqslant r$$

pentru variabila aleatoare X şi

$$(2)\qquad p_{*j} = \sum_{i=1}^{r} p_{ij} , \quad 1\leqslant j\leqslant s$$

pentru variabila aleatoare Y .

Mediile şi dispersiile variabilelor aleatoare X şi Y se definesc prin

$$M(X) = \sum_{i=1}^{r} x_i p_{i*} = \sum_{i=1}^{r} \sum_{j=1}^{s} x_i p_{ij}$$

$$D^2(X) = \sum_{i=1}^{r} [x_i - M(X)]^2 p_{i*} = \sum_{i=1}^{r} \sum_{j=1}^{s} [x_i - M(X)]^2 p_{ij}$$

$$M(Y) = \sum_{j=1}^{s} y_j p_{*j} = \sum_{i=1}^{r} \sum_{j=1}^{s} y_j p_{ij}$$

$$D^2(Y) = \sum_{j=1}^{s} [y_j - M(Y)]^2 p_{*j} = \sum_{i=1}^{r} \sum_{j=1}^{s} [y_j - M(Y)]^2 p_{ij}$$

Dacă V urmează o repartiţie probabilistică de tip continuu, atunci densităţile de repartiţie marginale se definesc prin

(3) $$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

pentru variabila aleatoare X şi

(4) $$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

pentru variabila aleatoare Y .

Mediile şi dispersiile variabilelor aleatoare X şi Y se definesc prin

$$M(X) = \int_{-\infty}^{+\infty} x\, f_1(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f(x,y)dx\, dy$$

$$D^2(X) = \int_{-\infty}^{+\infty} [x-M(X)]^2 f_1(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [x-M(X)]^2 f(x,y)dx\, dy$$

$$M(Y) = \int_{-\infty}^{+\infty} y\, f_2(y)dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f(x,y)dx\, dy$$

$$D^2(Y) = \int_{-\infty}^{+\infty} [y-M(Y)]^2 f_2(y)dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [y-M(Y)]^2 f(x,y)dx\, dy$$

Scanned with OKEN Scanner

Cu ajutorul probabilităţilor, respectiv densităţilor de repartiţie marginale se pot defini probabilităţile, respectiv densităţile de repartiţie condiţionate.

Astfel, probabilitatea evenimentului pentru care X ia valoarea x_i, condiţionată de evenimentul pentru care Y ia valoarea y_j, notată $p(x_i|y_j)$, se defineşte prin

$$p(x_i|y_j) = \frac{p_{ij}}{p_{*j}} . \tag{5}$$

Analog, probabilitatea evenimentului pentru care Y ia valoarea y_j, condiţionată de evenimentul pentru care X ia valoarea x_i, notată $p(y_j|x_i)$, se defineşte prin

$$p(y_j|x_i) = \frac{p_{ij}}{p_{i*}} . \tag{6}$$

Densitatea de repartiţie a lui X condiţionată de evenimentul pentru care Y ia valoarea y defineşte prin

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} . \tag{7}$$

Densitatea de repartiţie a lui Y condiţionată de evenimentul pentru care X ia valoarea x, se defineşte prin

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} . \tag{8}$$

Pentru j fixat, repartiţia probabilistică a lui X, condiţionată de evenimentul pentru care Y ia valoarea y_j este complet determinată de tabloul de repartiţie

$$X|Y = y_j : \begin{bmatrix} x_i \\ p(x_i|y_j) \end{bmatrix} \quad 1 \leq i \leq r$$

Intrucît j variază de la 1 la s, urmează că, variabilei aleatoare X i se pot asocia s repartiţii probabilistice condiţionate.

Analog, pentru i fixat, repartiţia probabilistică a lui Y,

condiţionată de evenimentul pentru care X ia valoarea x_i este complet determinată de tabloul de repartiţie

$$Y|X = x_i \; : \begin{bmatrix} y_j \\ p(y_j|x_i) \end{bmatrix} \quad 1 \leq j \leq s$$

Intrucît i variază de la 1 la r, urmează că, variabilei aleatoare Y i se pot asocia r repartiţii probabilistice condiţionate.

Mediile repartiţiilor condiţionate sînt date de

$$(9) \quad \begin{cases} M(X|Y = y_j) = \sum_{i=1}^{r} x_i p(x_i|y_j) = \frac{1}{p_{*j}} \sum_{i=1}^{r} x_i p_{ij} \;, & 1 \leq j \leq s \\ M(Y|X = x_i) = \sum_{j=1}^{s} y_j p(y_j|x_i) = \frac{1}{p_{i*}} \sum_{j=1}^{s} y_j p_{ij} \;, & 1 \leq i \leq r \end{cases}$$

dacă V are o repartiţie probabilistică de tip discret, şi de

$$(10) \quad \begin{cases} M(X|Y = y) = \int_{-\infty}^{+\infty} x \varphi(x|y)dx = \frac{1}{f_2(y)} \int_{-\infty}^{+\infty} x \, f(x,y)dx \\ M(Y|X = x) = \int_{-\infty}^{+\infty} y \, \phi(y\,x)dy = \frac{1}{f_1(x)} \int_{-\infty}^{+\infty} y \, f(x,y)dy \end{cases}$$

dacă V are o repartiţie probabilistică de tip continuu.

Dacă se notează

$$(11) \quad \begin{cases} M(X|Y = y_j) = \alpha_j \;, \; 1 \leq j \leq s \;, & M(X|Y = y) = \alpha(y) \\ M(Y|X = x_i) = \beta_i \;, \; 1 \leq i \leq r \;, & M(Y|X = x) = \beta(x) \end{cases}$$

atunci mărimile

$$M(X|Y) = \alpha(Y) \;, \quad M(Y|X) = \beta(X)$$

sînt variabile aleatoare a căror repartiţie probabilistică este complet determinată de tablourile de repartiţie

$$\alpha(Y) \; : \begin{bmatrix} \alpha_1 \cdots \alpha_j \cdots \alpha_s \\ p_{*1} \cdots p_{*j} \cdots p_{*s} \end{bmatrix} \;, \quad \beta(X) \; : \begin{bmatrix} \beta_1 \cdots \beta_i \cdots \beta_r \\ p_{1*} \cdots p_{i*} \cdots p_{r*} \end{bmatrix}$$

dacă V are o repartiţie probabilistică de tip discret şi de

$f_2(y)$ şi $f_1(x)$ dacă V are o repartiţie probabilistică de tip continuu.

Prin urmare, pentru variabilele aleatoare $\alpha(Y)$ şi $\beta(X)$ se pot calcula mediile şi dispersiile. Avem

$$(12)\begin{cases} M[\alpha(Y)]=\sum_{j=1}^{s}\alpha_j p_{*j}=\sum_{j=1}^{s}p_{*j}\,\frac{1}{p_{*j}}\sum_{i=1}^{r}x_i p_{ij}=\sum_{i=1}^{r}\sum_{j=1}^{s}x_i p_{ij}=M(X) \\ M[\beta(X)]=\sum_{i=1}^{r}\beta_i p_{i*}=\sum_{i=1}^{r}p_{i*}\,\frac{1}{p_{i*}}\sum_{j=1}^{s}y_j p_{ij}=\sum_{i=1}^{r}\sum_{j=1}^{s}y_j p_{ij}=M(Y) \end{cases}$$

Relaţiile (12) se mai pot scrie sub forma

$$(12')\qquad \begin{cases} M[M(X|Y)] = M(X) \\ M[M(Y|X)] = M(Y) \end{cases}$$

Dispersiile repartiţiilor condiţionate sînt date de

$$(13)\begin{cases} D^2(X|Y=y_j)=\sum_{i=1}^{r}[x_i-M(X|Y=y_j)]^2 p(x_i|y_j)= \\ = \frac{1}{p_{*j}}\sum_{i=1}^{r}[x_i-\alpha_j]^2 p_{ij} = d_j\ ,\ 1\leqslant j\leqslant s \\ D^2(Y|X=x_i) = \sum_{j=1}^{s}[y_j - M(Y|X=x_i)]^2 p(y_j|x_i) = \\ = \frac{1}{p_{i*}}\sum_{j=1}^{s}[y_j-\beta_i]^2 p_{ij} = b_i\ ,\ 1\leqslant i\leqslant r \end{cases}$$

dacă V are o repartiţie probabilistică de tip discret, şi de

$$(14)\begin{cases} D^2(X|Y=y) = \int_{-\infty}^{+\infty}[x-M(X|Y=y)]^2\varphi(x|y)dx = \\ = \frac{1}{f_2(y)}\int_{-\infty}^{+\infty}[x-\alpha(y)]^2 f(x,y)dx \\ D^2(Y|X=x) = \int_{-\infty}^{+\infty}[y-M(Y|X=x)]^2\phi(y|x)dy = \\ = \frac{1}{f_1(x)}\int_{-\infty}^{+\infty}[y-\beta(x)]^2 f(x,y)dy \end{cases}$$

dacă V are o repartiţie probabilistică de tip continuu.

Repartiţia variabilelor aleatoare $D^2(X|Y)$, $D^2(Y|X)$ este complet determinată de tablourile de repartiţie

$$D^2(X|Y) : \begin{bmatrix} d_1 \cdots d_j \cdots d_s \\ p_{*1} \cdots p_{*j} \cdots p_{*s} \end{bmatrix}, \quad D^2(Y|X) : \begin{bmatrix} b_1 \cdots b_i \cdots b_r \\ p_{1*} \cdots p_{i*} \cdots p_{r*} \end{bmatrix}$$

dacă V are o repartiţie probabilistică de tip discret şi de $\varphi(x|y), \psi(y|x)$ dacă V are o repartiţie probabilistică de tip continuu.

Vom demonstra că au loc relaţiile :

$$D^2(X) = M[X - \alpha(Y)]^2 + M[\alpha(Y) - M(X)]^2 \tag{15}$$

$$D^2(Y) = M[Y - \beta(X)]^2 + M[\beta(X) - M(Y)]^2 \tag{16}$$

Intr-adevăr, avem

$$D^2(X) = M[X-M(X)]^2 = M[(X-\alpha(Y))+(\alpha(Y)-M(X))]^2 = \tag{17}$$
$$= M[X - \alpha(Y)]^2 + M[\alpha(Y)-M(X)]^2 + 2M[(X-\alpha(Y))(\alpha(Y)-M(X))]$$

Intrucît

$$M[(X-\alpha(Y))(\alpha(Y)-M(X))] = [\alpha(Y)-M(X)][M(X)-M\alpha(Y)] =$$
$$= [\alpha(Y) - M(X)][M(X) - M(X)] = 0 ,$$

relaţia (17) devine tocmai relaţia (15). Analog se demonstrează relaţia (16). Ţinînd seama de relaţiile (14) şi (12'), relaţiile (15) şi (16) se mai pot scrie sub forma

$$D^2(X) = D^2(X|Y) + D^2[M(X|Y)] \tag{18}$$

$$D^2(Y) = D^2(Y|X) + D^2[M(Y|X)] \tag{19}$$

Definiţia 1. Se numeşte moment iniţial de ordin h în raport cu X şi de ordin k în raport cu Y, $(h,k = 0,1,2,\ldots)$, numărul

$$\alpha_{hk} = M(X^h Y^k)$$

dacă media din membrul drept există.

Evident, $\alpha_{10} = M(X)$, $\alpha_{01} = M(Y)$, $\alpha_{11} = M(XY)$.

Definiţia 2. Se numeşte moment centrat de ordin h în raport cu X şi de ordin k în raport cu Y, $(h,k = 0,1,2,\ldots)$, numărul

$$\mu_{hk} = M\left[(X - M(X))^h (Y - M(Y))^k\right]$$

dacă media din membrul drept există.

Evident,
$$\mu_{20} = M[X - M(X)]^2 = D^2(X)$$
$$\mu_{02} = M[Y - M(Y)]^2 = D^2(Y)$$
$$\mu_{11} = M\left[(X - M(X))(Y - M(Y))\right] = cov.(X,Y) =$$
$$= M(XY) - M(X)M(Y) = \alpha_{11} - \alpha_{10}\alpha_{01} .$$

Definiţia 3. Se spune că vectorul aleator bidimensional $V = (X,Y)$ urmează repartiţia normală de parametri $m_1, m_2, \sigma_1^2, \sigma_2^2, \rho$ dacă densitatea de repartiţie a lui este

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\frac{x-m_1}{\sigma_1}\frac{y-m_2}{\sigma_2} + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}$$

Prin calcul, se obţine :

$m_1 = M(X)$, $m_2 = M(Y)$, $\sigma_1^2 = D^2(X)$, $\sigma_2^2 = D^2(Y)$, ρ este coeficientul de corelaţie a variabilelor aleatoare X şi Y .

Teorema 1. Dacă vectorul $= (X,Y)$ urmează repartiţia normală de parametri $m_1, m_2, \sigma_1^2, \sigma_2^2, \rho$, atunci variabila aleatoare X urmează repartiţia $N(M_1, \sigma^2)$, iar variabila aleatoare Y urmează repartiţia $N(m_2, \sigma_2^2)$.

Demonstraţie. Conform relaţiei (3), densitatea de repartiţie corespunzătoare variabilei aleatoare X este

(20)
$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\frac{x-m_1}{\sigma_1}\frac{y-m_2}{\sigma_2} + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\} dy =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} \cdot$$

$$\cdot \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-m_2}{\sigma_2}\right)^2 - 2\rho\frac{x-m_1}{\sigma_1}\frac{y-m_2}{\sigma_2}\right]\right\} dy .$$

In urma schimbării de variabilă

(21) $$\frac{y - m_2}{\sigma_2} = t$$

relaţia (20) devine

(22) $$f_1(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} \cdot$$

$$\cdot\int_{-\infty}^{-\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[t^2 - 2\rho t\,\frac{x-m_1}{\sigma_1}\right]\right\}\sigma_2 dt =$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} \cdot$$

$$\cdot\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[t^2 - 2\rho t\,\frac{x-m_1}{\sigma_1} + \rho^2\left(\frac{x-m_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-m_1}{\sigma_1}\right)^2\right]\right\} dt =$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} \cdot$$

$$\cdot\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[t-\rho\frac{x-m_1}{\sigma_1}\right]^2 + \frac{\rho^2}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} dt =$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2 + \frac{\rho^2}{2(1-\rho^2)}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\} \cdot$$

$$\cdot\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[t - \rho\frac{x-m_1}{\sigma_1}\right]^2\right\} dt \quad .$$

In urma schimbării de variabilă

(23) $$\frac{t - \rho\dfrac{x - m_1}{\sigma_1}}{\sqrt{2}\sqrt{1-\rho^2}} = z$$

şi ţinînd seama că

$$\int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{\pi}$$

relaţia (22) devine

$$(24)\quad f_1(x)=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1-\rho^2}{2(1-\rho^2)}\cdot\left(\frac{x-m_1}{\sigma_1}\right)^2\right\}\sqrt{2}\int_{-\infty}^{+\infty}\sqrt{1-\rho^2}\,e^{-z^2}dz=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\}\int_{-\infty}^{+\infty}e^{-z^2}dz=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}$$

care este tocmai densitatea de repartiţie $N(m_1,\sigma_1^2)$. Printr-un calcul analog se găseşte

$$(25)\qquad f_2(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}.$$

§ 2. REPARTIŢII STATISTICE BIDIMENSIONALE.

Pînă acum am considerat că la indivizii unei colectivităţi se urmăreşte o singură caracteristică. In cele ce urmează vom considera că la indivizii unei colectivităţi se urmăresc simultan caracteristicile X şi Y. Dacă colectivitatea se interpretează ca un cîmp de evenimente, atunci X şi Y sînt componente ale unui vector aleator bidimensional (X,Y), numit caracteristică bidimensională a colectivităţii. Vom nota cu (x_i,y_i), $i\leq i\leq n$, o selecţie repetată de volum n şi cu (X_i,Y_i), $1\leq i\leq n$, vectorii aleatori de selecţie ai caracteristicii (X,Y).

Fie $x_1,\dots,x_i,\dots,x_r$ valorile distincte ale lui X şi $y_1,\dots,y_j,\dots,y_s$ valorile distincte ale lui Y, obţinute într-o selecţie repetată de volum n. O valoare a caracteristicii (X,Y) este (x_i,y_j).

Prin <u>frecvenţă absolută</u> a valorii (x_i,y_j) se înţelege numărul indivizilor din selecţia efectuată pentru care X ia valoarea x_i şi în acelaşi timp Y ia valoarea y_j. Dacă notăm frecvenţa absolută a valorii (x_i,y_j) cu n_{ij}, atunci evident

$\sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n$. Prin <u>frecvenţă relativă</u> a valorii (x_i,y_j) se înţelege raportul dintre frecvenţa absolută a acestei valori şi volumul selecţiei. Dacă notăm frecvenţa relativă a valorii (x_i,y_j) cu f_{ij}, deci $f_{ij} = \frac{n_{ij}}{n}$, atunci evident $\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = 1$.

Frecvenţele marginale absolute, respectiv relative, ale caracteristicii X, în selecţia efectuată, se definesc prin

$$(1) \qquad n_{i*} = \sum_{j=1}^{s} n_{ij} \ , \quad f_{i*} = \sum_{j=1}^{s} f_{ij} \ , \quad 1 \leqslant i \leqslant r \ .$$

Frecvenţele marginale absolute, respectiv relative, ale caracteristicii Y, în selecţia efectuată, se definesc prin

$$(2) \qquad n_{*j} = \sum_{i=1}^{r} n_{ij} \ , \ f_{*j} = \sum_{i=1}^{r} f_{ij} \ , \quad 1 \leqslant j \leqslant s \ .$$

Evident , $\sum_{i=1}^{r} n_{i*} = \sum_{j=1}^{s} n_{*j} = n$, $\sum_{i=1}^{r} f_{i*} = \sum_{j=1}^{s} f_{*j} = 1$.

Fiecare selecţie (x_i,y_i), $1 \leqslant i \leqslant n$, furnizează o repartiţie statistică bidimensională dată prin tabloul de repartiţie a frecvenţelor absolute

$$\left[\begin{matrix}(x_i,y_j)\\ n_{ij}\end{matrix}\right]_{\substack{1\leqslant i\leqslant r\\ 1\leqslant j\leqslant s}} \ , \quad \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n$$

sau a frecvenţelor relative

$$\left[\begin{matrix}(x_i,y_j)\\ f_{ij}\end{matrix}\right]_{\substack{1\leqslant i\leqslant r\\ 1\leqslant j\leqslant s}} \ , \quad \sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = 1$$

Tabloul de repartiţie al unei repartiţii statistice bidimensionale se numeşte <u>tablou de corelaţie</u> şi el se mai poate scrie sub forma tabelului 15

Tabelul 15

X \ Y	y_1	...	y_j	...	y_s	n_{i*}
x_1	n_{11}	...	n_{1j}	...	n_{1s}	n_{1*}
⋮	⋮		⋮		⋮	⋮
x_i	n_{i1}	...	n_{ij}	...	n_{is}	n_{i*}
⋮	⋮		⋮		⋮	⋮
x_r	n_{r1}	...	n_{rj}	...	n_{rs}	n_{r*}
n_{*j}	n_{*i}	.	n_{*j}	...	n_{*s}	n

Repartiţiile statistice marginale sînt date de tablourile de repartiţie

$$\begin{bmatrix} x_i \\ f_{i*} \end{bmatrix}_{1\leq i\leq r}, \quad \sum_{i=1}^{r} f_{i*} = 1$$

pentru caracteristica X , şi

$$\begin{bmatrix} y_j \\ f_{*j} \end{bmatrix}_{1\leq j\leq s}, \quad \sum_{j=1}^{s} f_{*j} = 1$$

pentru caracteristica Y. Media statistică a caracteristicii X se defineşte prin

$$(3) \qquad \bar{x} = \sum_{i=1}^{r} x_i f_{i*} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{s} x_i n_{ij} = \frac{1}{n} \sum_{i=1}^{r} x_i n_{i*}$$

iar media statistică a caracteristicii Y se defineşte prin

$$(4) \qquad \bar{y} = \sum_{j=1}^{s} y_j f_{*j} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{s} y_j n_{ij} = \frac{1}{n} \sum_{j=1}^{s} y_j n_{*j} .$$

Frecvenţele relative condiţionate se definesc prin

$$f_{i|j} = \frac{f_{ij}}{f_{*j}}$$

pentru caracteristica X, şi prin

$$f_{j|i} = \frac{f_{ij}}{f_{i*}}$$

pentru caracteristica Y .

Caracteristicii X i se pot asocia s repartiţii statistice condiţionate date de tablourile de repartiţie.

$$X|Y = y_j : \begin{bmatrix} x_1 & \cdots & x_i & \cdots & x_r \\ f_{1|j} & \cdots & f_{i|j} & \cdots & f_{r|j} \end{bmatrix} , \quad 1 \leqslant j \leqslant s$$

Caracteristicii Y i se pot asocia r repartiţii statistice condiţionate date de tablourile de repartiţie

$$Y|X = x_i : \begin{bmatrix} y_1 & \cdots & y_j & \cdots & y_s \\ f_{1|i} & \cdots & f_{j|i} & \cdots & f_{s|i} \end{bmatrix} , \quad 1 \leqslant i \leqslant r$$

Mediile statistice condiţionate sînt

$$\bar{M}(X|Y = y_j) = \alpha_j = \sum_{i=1}^{r} x_i f_{i|j} = \frac{1}{f_{*j}} \sum_{i=1}^{r} x_i f_{ij} , \quad 1 \leqslant j \leqslant s$$

$$M(Y|X = x_i) = \beta_i = \sum_{j=1}^{s} y_j f_{j|i} = \frac{1}{f_{i*}} \sum_{j=1}^{s} y_j f_{ij} , \quad 1 \leqslant i \leqslant r$$

Unind prin segmente de dreaptă punctele $P_j(y_j, \alpha_j)$, $1 \leqslant j \leqslant s$, se obţine <u>linia de regresie statistică a lui X în raport cu Y.</u> Analog, unind prin segmente de dreaptă punctele $Q_i(x_i, \beta_i)$, $1 \leqslant i \leqslant r$, se obţine <u>linia de regresie statistică a lui Y în raport cu X.</u>

Dispersia statistică a lui X se defineşte prin

$$(5)\qquad s_x^2 = \sum_{i=1}^{r}(x_i - \bar{x})^2 f_{i*} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \bar{x})^2 n_{ij} =$$

$$= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} x_i^2 n_{ij} - \bar{x}^2 = \frac{n\sum_{i=1}^{r} x_i^2 n_{i*} - (\sum_{i=1}^{r} x_i n_{i*})^2}{n^2}$$

Dispersia statistică a lui Y se defineşte prin

$$(6)\qquad s_y^2 = \sum_{j=1}^{s}(y_j - \bar{y})^2 f_{*j} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(y_j - \bar{y})^2 n_{ij} =$$

$$= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} y_j^2 n_{ij} - \bar{y}^2 = \frac{n\sum_{j=1}^{s} y_j^2 n_{*j} - (\sum_{j=1}^{s} y_j n_{*j})^2}{n^2}.$$

<u>Definiţia 1.</u> Se numeşte <u>moment iniţial statistic de ordin h în raport cu X şi de ordin k în raport cu Y</u>, $(h,k = 0,1,2,\ldots)$,

$$\bar{\alpha}_{hk} = \sum_{i=1}^{r}\sum_{j=1}^{s} x_i^h y_j^k f_{ij} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} x_i^h y_j^k n_{ij}$$

In particular, obţinem :

$$(7)\qquad \begin{cases} \bar{\alpha}_{10} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} x_i n_{ij} = \bar{x} \\ \bar{\alpha}_{01} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} y_j n_{ij} = \bar{y} \\ \bar{\alpha}_{11} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} x_i y_j n_{ij} \end{cases}$$

<u>Definiţia 2.</u> Se numeşte <u>moment centrat statistic de ordin h în raport cu X şi de ordin k în raport cu Y</u>, $(h,k = 0,1,2,\ldots)$, numărul

$$\bar{\mu}_{hk} = \sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \bar{x})^h (y_j - \bar{y})^k f_{ij} =$$

$$= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \bar{x})^h (y_j - \bar{y})^k n_{ij}.$$

In particular, obţinem :

$$(8)\quad \begin{cases} \bar{\mu}_{20} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i-\bar{x})^2 n_{ij} = s_x^2 \\ \bar{\mu}_{02} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(y_j-\bar{y})^2 n_{ij} = s_y^2 \\ \bar{\mu}_{11} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i-\bar{x})(y_j-\bar{y})n_{ij} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}x_i y_j n_{ij} - \overline{xy} = \\ = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}x_i y_j n_{ij} - (\frac{1}{n}\sum_{i=1}^{r}x_i n_{i*})(\frac{1}{n}\sum_{j=1}^{s}y_j n_{*j}) = \\ = \dfrac{n\sum_{i=1}^{r}\sum_{j=1}^{s}x_i y_j n_{ij} - (\sum_{i=1}^{r}x_i n_{i*})(\sum_{j=1}^{s}y_j n_{*j})}{n^2} \end{cases}$$

Definiţia 3. Se numeşte coeficient de corelaţie statistic a caracteristicilor X şi Y , numărul

$$\bar{r} = \frac{\bar{\mu}_{11}}{\sqrt{\bar{\mu}_{20}\bar{\mu}_{02}}}$$

Ţinînd seama de relaţiile (8), formula după care se calculează coeficientul de corelaţie statistic, pe baza tabloului de corelaţie, este

$$(9)\quad \bar{r} = \frac{n\sum_{i=1}^{r}\sum_{j=1}^{s}x_i y_j n_{ij} - (\sum_{i=1}^{r}x_i n_{i*})(\sum_{j=1}^{s}y_j n_{*j})}{\sqrt{\left[n\sum_{i=1}^{r}x_i^2 n_{i*} - (\sum_{i=1}^{r}x_i n_{i*})^2\right]\left[n\sum_{j=1}^{s}y_j^2 n_{*j} - (\sum_{j=1}^{s}y_j n_{*j})^2\right]}}$$

Dacă linia de regresie statistică a lui Y în raport cu X este o dreaptă, atunci ecuaţia ei este

$$y - \bar{y} = \bar{r}\sqrt{\frac{\bar{\mu}_{02}}{\bar{\mu}_{20}}}(x - \bar{x})$$

Dacă linia de regresie statistică a lui X în raport cu Y este o dreaptă, atunci ecuaţia ei este

$$x - \bar{x} = \bar{r}\sqrt{\frac{\bar{\mu}_{20}}{\bar{\mu}_{02}}}(y - \bar{y})$$

Se observă că dreptele de regresie statistică trec prin punctul $G(\bar{x},\bar{y})$. Mărimile

$$\bar{r}\sqrt{\frac{\bar{\mu}_{02}}{\bar{\mu}_{20}}} \quad , \quad \bar{r}\sqrt{\frac{\bar{\mu}_{20}}{\bar{\mu}_{02}}}$$

se numesc <u>coeficienţi de regresie statistică liniară</u>.

§ 3. COEFICIENTUL DE CORELAŢIE TEORETIC. COEFICIENTUL DE CONTINGENŢĂ.

Ţinînd seama de definiţiile 8 şi 9, §1, cap.I, coeficientul de corelaţie a variabilelor aleatoare X şi Y este dat de

$$(1) \qquad \rho(X,Y) = \frac{M[(X - M(X))(Y - M(Y))]}{D(X)D}$$

ceea ce este echivalent cu

$$(2) \qquad \rho(X,Y) = \frac{M(XY) - M(X)M(Y)}{D(X,Y)}$$

Coeficientul de corelaţie posedă următoarele proprietăţi:

1. $\rho(X,X) = 1$

Într-adevăr, folosind (2), se obţine

$$\rho(X,X) = \frac{M(X^2) - M^2(X)}{D^2(X)} = \frac{D^2(X)}{D^2(X)} = 1$$

2. $\rho(X,-X) = -1$

Într-adevăr, folosind (2), se obţine

$$\rho(X,-X) = \frac{M(-X^2)+M^2(X)}{D^2(X)} = -\frac{M(X^2)-M^2(X)}{D^2(X)} = -\frac{D^2(X)}{D^2(X)} = -1$$

3. Dacă ρ este coeficientul de corelaţie a variabilelor aleatoare X şi Y, iar ρ' este coeficientul de corelaţie a variabilelor aleatoare X' şi Y' , unde

$$X' = a(X - x_0) \ , \quad Y' = b(Y - y_0)$$

cu $a,b > 0$, $x_0 y_0 \in R$, atunci

$$\rho = \rho'$$

Intr-adevăr, dacă notăm

$$M(X) = m_1 \ , \quad M(Y) = m_2$$

avem

$$M(X') = a(m_1 - x_0) \ , \ M(Y') = b(m_2 - y_0)$$

$$D^2(X') = a^2 D^2(X) \ , \ D^2(Y') = b^2 D^2(Y) \ .$$

Atunci, folosind (1), se obţine

$$\rho' = \frac{M\{[a(X-x_0)-a(m_1-x_0)][b(Y-y_0)-b(m_2-y_0)]\}}{ab\, D(X)D(Y)} =$$

$$= \frac{abM[(X - m_1)(Y - m_2)]}{abD(X)D(Y)} = \rho .$$

4. Oricare ar fi variabilele aleatoare X şi Y, are loc

$$\rho^2(X,Y) \leqslant 1 \ .$$

Conform unei proprietăţi de la valoarea medie a unei variabile aleatoare, avem

$$|M(Z)| \leqslant M(|Z|)$$

oricare ar fi variabila aleatoare Z . Luînd

$$Z = (X - m_1)(Y - m_2)$$

se obţine

$$(3) \qquad \left|M[(X-m_1)(Y-m_2)]\right| \leqslant M[|X-m_1|\,|Y - m_2|]$$

Aplicînd lui (3), inegalitatea lui Schwarz , avem

$$(4) \qquad M[|X-m_1|\,|Y-m_2|] \leqslant \sqrt{M(X-m_1)^2 M(Y-m_2)^2}$$

Din relaţiile (3) şi (4), rezultă

$$(5) \qquad \left|M[(X-m_1)(Y-m_2)]\right| \leqslant \sqrt{M(X-m_1)^2 M(Y-m_2)^2}$$

ceea ce este echivalent cu

$$(6) \quad \left\{M[(X-m_1)(Y-m_2)]^2\right\} \leqslant M(X-m_1)^2 M(Y-m_2)^2 = D^2(X)D^2(Y)$$

de unde, rezultă

(7) $$\rho^2(X,Y) \leq 1 .$$

5. Dacă variabilele aleatoare X şi Y sînt independente, atunci $$\rho(X,Y) = 0 .$$

Intr-adevăr, folosind (2), rezultă

$$\rho(X,Y) = \frac{M(XY)-M(X)M(Y)}{D(X)D(Y)} = \frac{M(X)M(Y) - M(X)M(Y)}{D(X)D(Y)} = 0 .$$

Observaţia 1. Din $\rho(X,Y) = 0$ nu rezultă că X şi Y sînt independente. Se pot da exemple de variabile aleatoare X şi Y între care există o relaţie de dependenţă funcţională deşi coeficientul lor de corelaţie este nul.

6. Dacă între variabilele aleatoare X şi Y are loc relaţia $$Y = aX + b$$ unde $a,b \in R$, $a \neq 0$, atunci

$$\rho(X,Y) = \begin{cases} -1 & \text{dacă } a < 0 \\ 1 & \text{dacă } a > 0 \end{cases}$$

Intr-adevăr, dacă $m_1 = M(X)$, $m_2 = M(Y)$, atunci

$$m_2 = M(aX + b) = am_1 + b$$
$$D^2(Y) = D^2(aX + b) = a^2D^2(X)$$

Folosind (1), se obţine

$$\rho(X,Y) = \frac{M\left[(X-m_1)(aX+b-am_1-b)\right]}{\sqrt{D^2(X)a^2D^2(X)}} = \frac{aM\left[(X-m_1)^2\right]}{|a|D^2(X)} = \frac{aD^2(X)}{|a|D^2(X)} = \frac{a}{|a|} =$$

$$= \begin{cases} -1 & \text{dacă } a < 0 \\ 1 & \text{dacă } a > 0 \end{cases}$$

7. Dacă $\rho(X,Y) = \pm 1$, atunci între variabilele aleatoare X şi Y există o relaţie liniară.

Pentru demonstraţie a se vedea Mihoc Gh., Ciucu G., Craiu V. [10].

Teorema 1. Dacă vectorul aleator $V = (X,Y)$ urmează repartiţia normală, atunci din $\rho(X,Y) = 0$, rezultă că X şi Y sînt independente.

Demonstraţie. Făcînd $\rho = 0$ în expresia densităţii repartiţiei normale bidimensionale, se obţine

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\} =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{y-m_2}{\sigma_2}\right)^2}$$

ceea ce, conform teoremei 1, §1, cap. VII, se mai poate scrie sub forma $f(x,y) = f_1(x) f_2(y)$

unde $f_1(x)$ şi $f_2(y)$ sînt densităţile marginale de repartiţie. Ultima relaţie constituie tocmai o condiţie necesară şi suficientă de independenţă a componentelor X şi Y ale vectorului aleator V.

Dacă $\rho(X,Y) > 0$, se spune că între variabilele aleatoare X şi Y, există o corelaţie pozitivă (directă), ceea ce înseamnă că, cele două variabile aleatoare au acelaşi sens de variaţie. Dacă $\rho(X,Y) < 0$, se spune că între variabilele aleatoare X şi Y există o corelaţie negativă (inversă), ceea ce înseamnă că, cele două variabile aleatoare n-au acelaşi sens de variaţie. Dacă $\rho(X,Y)$ este apropiat de ± 1, atunci potrivit proprietăţii 7), între variabilele aleatoare X şi Y există o dependenţă funcţională liniară. Coeficientul de corelaţie a două variabile aleatoare, dă o indicaţie precisă asupra dependenţei funcţionale existente între ele, numai atunci cînd valoarea absolută a lui este apropiată de 1.

Fie X şi Y variabile aleatoare de tip discret date prin tablourile de repartiţie

$$X : \begin{bmatrix} x_i \\ p_{i*} \end{bmatrix}_{1 \leq i \leq r} , \quad Y : \begin{bmatrix} y_j \\ p_{*j} \end{bmatrix}_{1 \leq j \leq s}$$

unde p_{i*} şi p_{*j} sînt definiţi de formulele (1) şi (2) din § 1, cap.VII.

O indicaţie asupra dependenţei funcţionale existente între X şi Y este dată de <u>coeficientul de contingenţă</u>.

<u>Definiţia 1</u>. Se numeşte <u>coeficient de contingenţă</u> a variabilelor aleatoare X şi Y, numărul

$$K^2(X,Y) = \frac{1}{\sqrt{(r-1)(s-1)}} \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}p_{*j}}$$

Coeficientul de contingenţă posedă următoarele proprietăţi:

1. Dacă variabilele aleatoare X şi Y sînt independente, atunci $K^2(X,Y) = 0$

Intr-adevăr, din independenţa variabilelor aleatoare X şi Y rezultă

$$p_{ij} = p_{i*}p_{*j} , \quad 1 \leq i \leq r , \quad 1 \leq j \leq s$$

şi deci $K^2(X,Y) = 0$.

2. Dacă $K^2(X,Y) = 0$, atunci variabilele aleatoare X şi Y sînt independente.

Intr-adevăr, din $K^2(X,Y) = 0$, rezultă

$$p_{ij} = p_{i*}p_{*j} , \quad 1 \leq i \leq r , \quad 1 \leq j \leq s$$

ceea ce înseamnă că X şi Y sînt independente.

3. Oricare ar fi variabilele aleatoare X şi Y are loc

$$0 \leq K^2(X,Y) \leq 1 .$$

Pentru demonstraţie a se vedea Onicescu O, Mihoc Gh. [12] .

4. Dacă între variabilele aleatoare X şi Y există o relaţie de dependenţă funcţională, atunci $K^2(X,Y) = 1$.

Pentru demonstraţie a se vedea Onicescu O., Mihoc Gh. [12].

Observaţia 2. Proprietatea reciprocă a proprietăţii 4) nu este adevărată.

Observaţia 3. In demonstrarea proprietăţilor 1 şi 2 am ţinut seama de faptul că, o condiţie necesară şi suficientă pentru ca variabilele aleatoare de tip discret X şi Y să fie independente este

$$p_{ij} = p_{i*}p_{*j} \quad , \quad 1 \leq i \leq r \quad , \quad 1 \leq j \leq s.$$

§ 4. REGRESIA SIMPLĂ TEORETICĂ .

Definiţia 1. Se numeşte regresie a lui X în raport cu Y variabila aleatoare $\alpha(Y) = M(X|Y)$.

Definiţia 2. Se numeşte regresie a lui Y în raport cu X variabila aleatoare $\beta(X) = M(Y|X)$.

Definiţia 3. Se numeşte regresie simplă regresia unei variabile aleatoare în raport cu o singură variabilă aleatoare.

Definiţia 4. Se spune că între două variabile aleatoare există o dependenţă stohastică, atunci cînd la variaţia uneia corespunde variaţia repartiţiei probabilistice a celeilalte.

Ţinînd seama de definiţiile precedente, putem afirma că , regresia simplă este o formă a dependenţei stohastice dintre două variabile aleatoare.

Definiţia 5. Se spune că regresia lui X în raport cu Y este liniară dacă $M(X|Y = y) = a'y + b'$.

Definiţia 6. Se spune că regresia lui Y în raport cu X este liniară, dacă $M(Y|X = x) = ax + b$.

Observaţie. Dacă una dintre regresii este liniară, nu rezultă că şi cealaltă regresie este liniară.

Teorema 1. Dacă vectorul aleator $V = (X,Y)$ urmează o repartiţie normală, atunci ambele regresii sînt liniare.

Demonstraţie. Potrivit relaţiei (8) şi teoremei 1 din §1, cap.VII, densitatea de repartiţie a lui Y, condiţionată de evenimentul pentru care X ia valoarea x este

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{y-m_2}{\sigma_2} - \rho\frac{x-m_1}{1}\right]^2\right\} =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y-m_2-\rho\frac{\sigma_2}{\sigma_1}(x-m_1)\right]^2\right\}.$$

Din relaţia ce exprimă pe $\psi(y|x)$ se observă că

$$M(Y|X = x) = \beta(x) = m_2 + \rho\frac{\sigma_2}{\sigma_1}(x - m_1)$$

ceea ce înseamnă, conform definiţiei 6 că, regresia lui Y în raport cu X este liniară.

Printr-un calcul analog se găseşte că, densitatea de repartiţie a lui X condiţionată de evenimentul pentru care Y ia valoarea y , este

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x-m_1}{\sigma_1} - \rho\frac{y-m_2}{\sigma_2}\right]^2\right\} =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x - m_1 - \rho\frac{y-m_2}{\sigma_2}\right]^2\right\}$$

Din expresia lui $\varphi(x|y)$ se observă că

$$M(X|Y = y) = m_1 + \rho\frac{\sigma_1}{\sigma_2}(y - m_2)$$

ceea ce înseamnă, conform definiției 5 că, regresia lui X asupra lui Y este liniară.

Teorema 2. Dacă una din regresiile vectorului aleator V = (X,Y) este o constantă, atunci

$$cov.(X,Y) = 0 .$$

Demonstrație. Fie V = X,Y) un vector aleator urmînd o repartiție probabilistică de tip continuu cu densitatea de repartiție f(x,y). Presupunem că regresia lui Y în raport cu X este constantă, deci, M(Y|X = x) = c. Valoarea c poate fi determinată. Intr-adevăr, avem

$$(1) \qquad c = M(Y|X = x) = \int_{-\infty}^{+\infty} y\,\psi(y|x)dy =$$

$$= \int_{-\infty}^{+\infty} y\,\frac{f(x,y)}{f_1(x)}\,dy = \frac{1}{f_1(x)}\int_{-\infty}^{+\infty} yf(x,y)dy$$

sau

$$(2) \qquad cf_1(x) = \int_{-\infty}^{+\infty} yf(x,y)dy$$

Integrînd relația (2), în raport cu x de la $-\infty$ la $+\infty$, se obține

$$(3) \qquad c\int_{-\infty}^{+\infty} f_1(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\,f(x,y)dx\,dy$$

sau

$$(4) \qquad c = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dx\,dy = M(Y)$$

Conform definiției 8, §1, cap.I, avem

$$cov.(X,Y)=M(XY)-M(X)M(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - cM(X) =$$

$$=\int_{-\infty}^{+\infty} xf_1(x)dx\int_{-\infty}^{+\infty}\frac{f(x,y)}{f_1(x)}\,dy-cM(X)=\int_{-\infty}^{+\infty} xf_1(x)dx\int_{-\infty}^{+\infty} y\,\psi(y|x)dy-cM(X)=$$

$= M(X)M(Y|X = x) - cM(X) = cM(X) - cM(X) = 0$.

Intrucît, conform definiţiei 9, §1, cap.I, avem

$$\rho(X,Y) = \frac{cov.(X,Y)}{D(X)D(Y)}$$

rezultă că

$$\rho(X,Y) = 0 .$$

§ 5. REGRESIA TEORETICĂ LINIARĂ.

Fie $f(x,y)$ densitatea de repartiţie comună a variabilelor aleatoare X şi Y, pentru care există şi sînt finite

$$m_1 = M(X),\ m_2 = M(Y),\ \sigma_1^2 = D^2(X),\ \sigma_2^2 = D^2(Y) \text{ şi } \rho(X,Y).$$

Presupunem că regresia lui Y în raport cu X este liniară. Din definiţia 6, §3, cap.VII, rezultă

(1) $$M(Y|X = x) = ax + b .$$

Ne propunem să determinăm coeficienţii a şi b în funcţie de $m_1, m_2, \sigma_1^2, \sigma_2^2, \rho$. Folosind relaţiile (10), §1, cap.VII, se obţine

(2) $$M(Y|X = x) = \frac{1}{f_1(x)} \int_{-\infty}^{+\infty} yf(x,y)dy .$$

Din (1) şi (2), rezultă

(3) $$\int_{-\infty}^{+\infty} yf(x,y)dy = axf_1(x) + bf_1(x) .$$

Integrînd relaţia (3), în raport cu x de la $-\infty$ la $+\infty$, se obţine

(4) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy = a\int_{-\infty}^{+\infty} xf_1(x)dx + b\int_{-\infty}^{+\infty} f_1(x)dx$$

sau

(5) $$M(Y) = aM(X) + b .$$

Ţinînd seama de notaţiile de mai înainte, relaţia (5) se mai poate scrie sub forma

(6) $$m_2 = am_1 + b .$$

Inmulţind relaţia (3) cu x şi integrînd rezultatul obţinut în raport cu x de la $-\infty$ la $+\infty$, rezultă

$$(7) \qquad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = a\int_{-\infty}^{+\infty} x^2 f_1(x)dx + b\int_{-\infty}^{+\infty} xf_1(x)dx$$

sau

$$(8) \qquad M(XY) = aM(X^2) + bM(X) .$$

Rezolvînd, în raport cu a şi b, sistemul de ecuaţii determinat de relaţiile (6) şi (8), se obţine

$$(9) \qquad a = \rho\frac{\sigma_2}{\sigma_1} \quad , \quad b = m_2 - \rho\frac{\sigma_2}{\sigma_1} m_1 .$$

Inlocuind a şi b daţi de (9), în relaţia (1), se obţine

$$(10) \qquad y - m_2 = \rho\frac{\sigma_2}{\sigma_1}(x - m_1)$$

ceea ce reprezintă ecuaţia dreptei de regresie teoretică a lui Y în raport cu X .

Dacă regresia lui X în raport cu Y este liniară, ceea ce conform definiţiei 5 înseamnă că are loc

$$(11) \qquad M(X|Y = y) = a'y + b'$$

atunci printr-un calcul analog se găseşte

$$(12) \qquad a' = \rho\frac{\sigma_1}{\sigma_2} \quad , \quad b' = m_1 - \rho\frac{\sigma_1}{\sigma_2} m_2 ,$$

Inlocuind a' şi b' daţi de (12) în relaţia (11), se obţine

$$(13) \qquad x - m_1 = \rho\frac{\sigma_1}{\sigma_2}(y - m_2)$$

ceea ce este echivalent cu

$$(14) \qquad y - m_2 = \frac{\sigma_2}{\rho\sigma_1}(x - m_1) .$$

Coeficienţii a şi a' se numesc <u>coeficienţi de regresie teoretică liniară.</u> Din (9) şi (12) se observă că, a şi a' au

același semn cu coeficientul de corelație teoretic ρ. Din (10) și (14) se observă că, dreptele de regresie trec prin punctul $G(m_1, m_2)$.

Făcînd $\rho = 0$ în (10) și (14), rezultă

$$x = m_1 \ , \quad y = m_2$$

ceea ce înseamnă că dreptele de regresie sînt paralele cu axele de coordonate. In cazul $\rho = 1$ ecuațiile (10) și (14) sînt identice, deci, dreptele de regresie se confundă. Din (9) și (12), rezultă

$$aa' = \rho^2 .$$

Legătura dintre coeficienții dreptei de regresie și o problemă de minim este dată de

Teorema 1. Minimul valorii medii

$$M\left[Y - \lambda X - \mu\right]^2$$

cu $\lambda, \mu \in R$ este atins pentru $\lambda = a$ și $\mu = b$, unde a și b sînt dați de (9), iar acest minim este egal cu $\sigma_2^2(1-\rho^2)$.

Demonstrație. Presupunem că vectorul aleator $V = (X,Y)$ urmează o repartiție probabilistică de tip discret, dată de tabloul de repartiție

$$\begin{bmatrix} (x_i, y_j) \\ p_{ij} \end{bmatrix}_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}}$$

Dacă se notează

$$F(\lambda, \mu) = M\left[Y - \lambda X - \mu\right]^2$$

atunci

$$F(\lambda, \mu) = \sum_{i=1}^{r} \sum_{j=1}^{s} (y_j - \lambda x_i - \mu)^2 p_{ij} .$$

Valorile (λ, μ) ce realizează minimul lui $F(\lambda, \mu)$, în cazul cînd acest minim există, se află printre soluțiile sistemului de ecuații

$$(15)\quad \begin{cases} \dfrac{\partial F(\lambda,\mu)}{\partial \lambda} = -2\sum_{i=1}^{r}\sum_{j=1}^{s} x_i(y_j - \lambda x_i - \mu)p_{ij} = 0 \\ \dfrac{\partial F(\lambda,\mu)}{\partial \mu} = -2\sum_{i=1}^{r}\sum_{j=1}^{s} (y_j - \lambda x_i - \mu)p_{ij} = 0 \end{cases}$$

care se mai poate scrie sub forma

$$(16)\quad \begin{cases} \lambda\sum_{i=1}^{r}\sum_{j=1}^{s} x_i^2 p_{ij} + \mu\sum_{i=1}^{r}\sum_{j=1}^{s} x_i p_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{s} x_i y_j p_{ij} \\ \lambda\sum_{i=1}^{r}\sum_{j=1}^{s} x_i p_{ij} + \mu\sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{s} y_j p_{ij} \end{cases}$$

sau

$$(17)\quad \begin{cases} \lambda M(X^2) + \mu M(X) = M(XY) \\ \lambda M(X) + \mu = M(Y) \end{cases}$$

Rezolvînd sistemul de ecuaţii (17), se obţine

$$(18)\quad \lambda = \rho\frac{\sigma_2}{\sigma_1}\ ,\quad \mu = m_2 - \rho\frac{\sigma_2}{\sigma_1}\,m_1$$

ceea ce pe baza lui (9), revine la

$$(19)\quad \lambda = a\ ,\quad \mu = b$$

Vom arăta că

$$\min_{\lambda,\mu} F(\lambda,\mu) = F(a,b)\,.$$

Pentru a avea loc ultima relaţie, trebuie îndeplinite condiţiile

$$(20)\quad \begin{cases} \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda^2}\cdot\dfrac{\partial^2 F(\lambda,\mu)}{\partial\mu^2} - \left(\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda\,\partial\mu}\right)^2\right]_{(a,b)} > 0 \\ \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda^2}\right]_{(a,b)} > 0 \quad \text{sau} \quad \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\mu^2}\right]_{(a,b)} > 0 \end{cases}$$

Ţinînd seama de (15), rezultă

$$(21)\quad \begin{cases} \dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda^2} = 2\sum_{i=1}^{r}\sum_{j=1}^{s} x_i^2 p_{ij} = 2M(X^2) \\ \dfrac{\partial^2 F(\lambda,\mu)}{\partial\mu^2} = 2\sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} = 2 \\ \dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda\partial\mu} = 2\sum_{i=1}^{r}\sum_{j=1}^{s} x_i p_{ij} = 2M(X) \end{cases}$$

Folosind (21), se obţine

$$(22)\quad \begin{cases} \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda^2}\cdot\dfrac{\partial^2 F(\lambda,\mu)}{\partial\mu^2} - \left(\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda\,\partial\mu}\right)^2\right]_{(a,b)} = 4D^2(X) > 0 \\ \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\lambda^2}\right]_{(a,b)} = 2M(X^2) > 0 \\ \left[\dfrac{\partial^2 F(\lambda,\mu)}{\partial\mu^2}\right]_{(a,b)} = 4 > 0 \end{cases}$$

Pe baza lui (22), putem afirma că punctul (a,b) este un punct de minim pentru $F(\lambda,\mu)$.

Calculînd valoarea lui $F(\lambda,\mu)$ în punctul (a,b) se obţine

$$(23)\quad F(a,b) = M\left[Y - \rho\frac{\sigma_2}{\sigma_1}X - m_2 + \rho\frac{\sigma_2}{\sigma_1}m_1\right]^2 =$$

$$= M\left[(Y - m_2) - \rho\frac{\sigma_2}{\sigma_1}(X - m_1)\right]^2 =$$

$$= M\left[(Y-m_2)^2 + \rho^2\frac{\sigma_2^2}{\sigma_1^2}(X-m_1)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(X - m_1)(Y - m_2)\right] =$$

$$= M\left[(Y-m_2)^2\right] + \rho^2\frac{\sigma_2^2}{\sigma_1^2}M\left[(X-m_1)^2\right] - 2\rho\frac{\sigma_2}{\sigma_1}M\left[(X-m_1)(Y-m_2)\right] =$$

$$= \sigma_2^2 + \rho^2\frac{\sigma_2^2}{\sigma_1^2}\cdot\sigma_1^2 - 2\rho\frac{\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2 = \sigma_2^2(1 - \rho^2)\,.$$

Cu aceasta teorema este demonstrată.

Analog, se poate demonstra

Teorema 2. Minimul valorii medii

$$M[X - \lambda' Y - \mu']^2$$

cu $\lambda', \mu' \in R$ este atins pentru $\lambda' = a'$ şi $\mu' = b'$, unde a' şi b' sînt daţi de (12), iar acest minim este egal cu $\sigma_1^2(1 - \rho^2)$.

Observaţie. Teoremele 1 şi 2 justifică folosirea metodei celor mai mici pătrate pentru determinarea coeficienţilor a, b, a', b' din regresiile simple liniare.

§ 6. RAPORTUL DE CORELAŢIE TEORETIC .

Ţinînd seama de relaţiile (12'), §1, cap.VII, relaţiile (15) şi (16), §1, cap.VII, devin

$$(1) \qquad D^2(X) = M[X - \alpha(Y)]^2 + D^2[\alpha(Y)]$$

$$(2) \qquad D^2(Y) = M[Y - \beta(X)]^2 + D^2[\beta(X)]$$

Definiţia 1. Se numeşte raport de corelaţie a variabilei aleatoare X în raport cu variabila aleatoare Y , numărul

$$(3) \qquad \theta^2_{x|y} = \frac{D^2[\alpha(Y)]}{D^2(X)}$$

Definiţia 2. Se numeşte raport de corelaţie a variabilei aleatoare Y în raport cu variabila aleatoare X, numărul

$$(4) \qquad \theta^2_{y|x} = \frac{D^2[\beta(X)]}{D^2(Y)}$$

Ţinînd seama de (1), relaţia (3), devine

$$(5) \qquad \theta^2_{x|y} = 1 - \frac{M[X - \alpha(Y)]^2}{D^2(X)}$$

de unde rezultă

$$(6) \qquad 1 - \theta^2_{x|y} = \frac{M[X - \alpha(Y)]^2}{D^2(X)} \quad .$$

Ținînd seama de (2), relația (4), devine

$$\theta^2_{y|x} = 1 - \frac{M[Y - \beta(X)]^2}{D^2(Y)} \tag{7}$$

de unde rezultă

$$1 - \theta^2_{y|x} = \frac{M[Y - \beta(X)]^2}{D^2(Y)} \tag{8}$$

Vom demonstra că are loc relația

$$M[Y - aX - b]^2 = M[Y - \beta(X)]^2 + M[\beta(X) - aX - b]^2 . \tag{9}$$

Intr-adevăr, avem

$$\begin{aligned} M[Y - aX - b]^2 &= M[(Y - \beta(X)) + (\beta(X) - aX - b)]^2 = \\ &= M[Y - \beta(X)]^2 + M[\beta(X) - aX - b]^2 + \\ &+ 2M[(Y - \beta(X))(\beta(X) - aX - b)]. \end{aligned} \tag{10}$$

Ținînd seama de relațiile (12'), §1, cap.VII, rezultă

$$\begin{aligned} M\{[Y - \beta(X)][\beta(X) - aX - b]\} &= \\ = [\beta(X) - aX - b][M(Y) - M(\beta(X))] &= \\ = [\beta(X) - aX - b][M(Y) - M(Y)] &= 0 . \end{aligned} \tag{11}$$

Avînd în vedere (11), relația (10), devine tocmai relația (9). Conform teoremei 1, §5, cap.VII, pentru

$$a = \rho\frac{\sigma_2}{\sigma_1}, \quad b = m_2 - \rho\frac{\sigma_2}{\sigma_1}m_1$$

expresia

$$M[Y - aX - b]^2$$

își atinge minimul egal cu

$$\sigma_2^2(1 - \rho^2) .$$

In acest caz, relația (9) devine

$$\sigma_2^2(1 - \rho^2) = M[Y - \beta(X)]^2 + M[\beta(X) - aX - b]^2 \tag{12}$$

ceea ce este echivalent cu

$$1 - \rho^2 = \frac{M[Y - \beta(X)]^2}{\sigma_2^2} + \frac{M[\beta(X) - aX - b]^2}{\sigma_2^2} \tag{13}$$

Tinînd seama de (8), relaţia (13) devine

$$(14) \qquad 1 - \rho^2 = 1 - \theta^2_{y|x} + \frac{M\left[\beta(X) - aX - b\right]^2}{\sigma_2^2}$$

ceea ce este echivalent cu

$$(15) \qquad \theta^2_{y|x} = \rho^2 + \frac{M\left[\beta(X) - aX - b\right]^2}{\sigma_2^2}$$

Raportul de corelaţie posedă următoarele proprietăţi:

1. Oricare ar fi variabilele aleatoare X şi Y , are loc

$$0 \leqslant \theta^2_{y|x} \leqslant 1$$

Intr-adevăr, faptul că $\theta^2_{y|x} \geqslant 0$ rezultă din (4). Din relaţia (8), rezultă $1 - \theta^2_{y|x} \geqslant 0$, adică, $\theta^2_{y|x} \leqslant 1$.

2. $\theta^2_{y|x} = 0$, dacă şi numai dacă $\beta(X)$ este o constantă.

Intr-adevăr, dacă $\theta^2_{y|x} = 0$, atunci din (4) rezultă $D^2\left[\beta(X)\right] = 0$, adică, $\beta(X)$ este o constantă. Dacă $\beta(X)$ este o constantă, atunci conform teoremei 2, §4, cap.VII, rezultă $\rho = 0$, $a = 0$, $b = m_2$. Tinînd seama de toate acestea, din relaţia (15), se obţine $\theta^2_{y|x} = 0$.

3. Dacă $\theta^2_{y|x} = 1$, atunci între variabilele aleatoare X şi Y există o dependenţă funcţională.

Intr-adevăr, dacă $\theta^2_{y|x} = 1$, atunci din relaţia (8), rezultă

$$M\left[Y - \beta(X)\right]^2 = 0$$

ceea ce este echivalent cu

$$Y - \beta(X) = 0$$

Aceleaşi proprietăţi de mai înainte le posedă şi $\theta^2_{x|y}$. Cu ajutorul raportului de corelaţie se apreciază intensitatea corelaţiei neliniare .

§ 7. COEFICIENTUL DE CORELAȚIE DE SELECȚIE.

Fie (X,Y) o caracteristică bidimensională urmărită la indivizii unei colectivități și (X_i,Y_i), $1 \leqslant i \leqslant n$, vectorii aleatori de selecție.

Definiția 1. Se numește coeficient de corelație de selecție statistica

$$r(X,Y;n) = \frac{\sum_{i=1}^{n}[X_i - m(X;n)][Y_i - m(Y;n)]}{\sqrt{\sum_{i=1}^{n}[X_i - m(X;n)]^2 \sum_{i=1}^{n}[Y_i - m(Y;n)]^2}}$$

unde

$$m(X;n) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , \quad m(Y;n) = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad .$$

Teorema 1. Oricare ar fi caracteristica bidimensională (X,Y) a unei colectivități, are loc

$$r^2(X,Y;n) \leqslant 1 \quad .$$

Demonstrație. Evident, oricare ar fi $a \in R$, are loc

$$(1) \qquad \sum_{i=1}^{n}[a(X_i - m(X;n)) + (Y_i - m(Y;n))]^2 \geqslant 0$$

ceea ce este echivalent cu

$$(2) \qquad a^2 \sum_{i=1}^{n}[X_i - m\,X;n)]^2 + 2a \sum_{i=1}^{n}[X_i - m(X;n)][Y_i - m(Y;n)] +$$

$$+ \sum_{i=1}^{n}[Y_i - m(Y;n)]^2 \geqslant 0 \quad .$$

Membrul stîng al inegalității (2) reprezintă o formă pătratică în „a" . Pe baza lui (2), urmează că , discriminantul acestei forme pătratice trebuie să fie nenegativ, ceea ce revine la

$$(3) \qquad \sum_{i=1}^{n}[X_i - m(X;n)]^2 \sum_{i=1}^{n}[Y_i - m(Y;n)]^2 -$$

$$-\left\{\sum_{i=1}^{n}[X_i - m(X;n)][Y_i - m(Y;n)]\right\}^2 \geqslant 0 .$$

Inegalitatea (3) este echivalentă cu inegalitatea

$$(4)\qquad \left\{\sum_{i=1}^{n}[X_i - m(X;n)][Y_i - m(Y;n)]\right\}^2 \leqslant$$

$$\leqslant \sum_{i=1}^{n}[X_i - m(X;n)]^2 \sum_{i=1}^{n}[Y_i - m(Y;n)]^2$$

Ținînd seama de definiția coeficientului de corelație de selecție din relația (4), rezultă

$$(5)\qquad r^2(X,Y;n) \leqslant 1 .$$

Observație. Se poate demonstra că, coeficientul de corelație de selecție $r(X,Y;n)$, este o funcție de estimație consistentă pentru coeficientul de corelație teoretic ϱ .

Fie (X,Y) o caracteristică bidimensională urmărită la indivizii unei colectivități, avînd o repartiție normală de parametri $m_1 = m_2 = 0$, $\sigma_1^2, \sigma_2^2, \varrho$.

Teorema 2. Dacă $\varrho(X,Y) = 0$, atunci statistica

$$\sqrt{n-1}\, r(X,Y;n)$$

urmează asimptotic repartiția $N(0,1)$.

Demonstrație. A se vedea Iosifescu M., Mihoc Gh., Teodorescu R. [7] .

Folosind coeficientul de corelație de selecție $r(X,Y;n)$, R.A, Fisher a considerat statistica

$$Z(X,Y;n) = \frac{1}{2}\ln\frac{1 + r(X,Y;n)}{1 - r(X,Y;n)}$$

care urmează, chiar pentru valori mici ale lui n, aproximativ repartiția normală, cu

$$M(Z) \simeq \frac{1}{2}\ln\frac{1+\varrho}{1-\varrho} + \frac{\varrho}{2(n-1)}$$

$$D^2(Z) \simeq \frac{1}{n-3}$$

Statistica $Z(X,Y;n)$ va fi folosită pentru :

I. construirea unui test de verificare a ipotezei

$$H_o : \rho = \rho_o$$

II. construirea unui test de verificare a ipotezei

$$H_1 : \rho_1 = \rho_2$$

III. construirea unui interval de încredere pentru ρ .

In cele ce urmează prezentăm rezolvarea celor trei probleme formulate .

I. Fie (X,Y) o caracteristică bidimensională urmărită la indivizii unei colectivităţi avînd o repartiţie normală de parametri $m_1, m_2, \sigma_1^2, \sigma_2^2, \rho$. Fie ρ_o un număr real şi $-1 < \rho_o < 1$. Din colectivitate se efectuează selecţia repetată (x_i, y_i), $1 < i < n$. Pe baza acestei selecţii trebuie testată ipoteza

$$H_o \quad : \rho_o .$$

Testul se bazează pe criteriul

$$C(X,Y;n) = \left[Z(X,Y;n) - \frac{1}{2} \ln \frac{1+\rho}{1-\rho}\right] \sqrt{n-3}$$

Dacă ipoteza H_o se acceptă, atunci $C(X,Y;n)$ devine

$$U(X,Y;n) = \left[Z(X,Y;n) - \frac{1}{n} \ln \frac{1+\rho_o}{1-\rho_o}\right] \sqrt{n-3}$$

care, conform teoremei 6, §3, cap.I, urmează aproximativ repartiţia $N(0,1)$.

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea u_α, astfel încît să aibă loc

$$P^{(n)}\left(\omega^{(n)} \in \Omega^{(n)} : |U(X,Y;n)| < u_\alpha\right) = 1 - \alpha .$$

Valoarea critică u_α se determină, cu ajutorul tabelului II din anexă, pe baza relaţiei

$$\Phi(u_\alpha) = \frac{1-\alpha}{2}$$

Valoarea statisticii $U(X,Y;n)$, pentru selecţia efectuată, est

$$U_o = \left[Z_o - \frac{1}{2}\ln\frac{1+\rho_o}{1-\rho_o}\right]\sqrt{n-3}$$

unde

$$Z_o = \frac{1}{2}\ln\frac{1+\bar{r}}{1-\bar{r}}$$

Regiunea critică a testului este

$$U = \left\{(x_i,y_i) : 1 \leq i \leq n : \left|\left(Z_o - \frac{1}{2}\ln\frac{1+\rho_o}{1-\rho_o}\right)\sqrt{n-3}\right| > u_\alpha\right\}$$

iar mulţimea din R , corespunzătoare lui U este

$$I = (-\infty, -u_\alpha) \cup (u_\alpha, +\infty).$$

Testul este :

i) dacă $U_o \in [-u_\alpha, u_\alpha]$, ipoteza H_o se acceptă

ii) dacă $U_o \notin [-u_\alpha, u_\alpha]$, ipoteza H_o se respinge

II. Fie (X_1,Y_1) caracteristica urmărită la indivizii colectivităţii C_1 avînd o repartiţie normală cu coeficientul de corelaţie ρ_1 şi (X_2,Y_2) caracteristica urmărită la indivizii colectivităţii C_2 avînd o repartiţie normală cu coeficientul de corelaţie ρ_2. Fie $(x_i^{(1)},y_i^{(1)})$, $1 \leq i \leq n_1$, o selecţie repetată din C_1 şi $(x_j^{(2)},y_j^{(2)})$, $1 \leq j \leq n_2$, o selecţie repetată din C_2. Pe baza acestor selecţii trebuie testată ipoteza

$$H_1 : \rho_1 = \rho_2$$

Notăm

$$Z(X_1,Y_1;n_1) = \frac{1}{2}\ln\frac{1+r(X_1,Y_1;n_1)}{1-r(X_1,Y_1;n_1)}\ , \quad Z(X_2,Y_2;n_2) = \frac{1}{2}\ln\frac{1+r(X_2,Y_2;n_2)}{1-r(X_2,Y_2;n_2)}$$

Testul se bazează pe criteriul

$$C(X_1,Y_1,X_2,Y_2;n_1,n_2) = \frac{\left[Z(X_1,Y_1;n_1) - \frac{1}{2}\ln\frac{1+\rho_1}{1-\rho_1}\right] - \left[Z(X_2,Y_2;n_2) - \frac{1}{2}\ln\frac{1+\rho_2}{1-\rho_2}\right]}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

Dacă ipoteza H_1 se acceptă, atunci $C(X_1,Y_1,X_2;n_1,n_2)$ devine

$$U(X_1,Y_1,X_2,Y_2;n_1,n_2) = \frac{Z(X_1,Y_1;n_1) - Z(X_2,Y_2;n_2)}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

care, conform teoremei 6, §3, cap.I, urmează aproximativ repartiţia N(0,1).

Pentru nivelul de semnificaţie α fixat, se poate determina valoarea u_α astfel încît să aibă loc

$$P^{(n)}(\omega^{(n)} \in \Omega^{(n)}; |U(X_1,Y_1,X_2,Y_2;n_1,n_2)| < u_\alpha) = 1 - \alpha .$$

Valoarea critică u_α se determină, cu ajutorul tabelului IV din anexă pe baza relaţiei

$$\Phi(u_\alpha) = \frac{1-\alpha}{2}$$

Valoarea statisticii $U(X_1,Y_1,X_2,Y_2;n_1,n_2)$ pentru selecţiile efectuate este

$$U_0 = \frac{\bar{Z}_1 - \bar{Z}_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

unde

$$\bar{Z}_1 = \frac{1}{2}\ln\frac{1+\bar{r}_1}{1-\bar{r}_1} \quad , \quad \bar{Z}_2 = \frac{1}{2}\ln\frac{1+\bar{r}_2}{1-\bar{r}_2}$$

Regiunea critică a testului este

$$U = \left\{(x_i^{(1)},y_i^{(1)}), 1 \le i \le n_1; (x_j^{(2)},y_j^{(2)}), 1 \le j \le n_2 : \left|\frac{\bar{Z}_1 - \bar{Z}_2}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}}\right| > u_\alpha\right\}$$

iar mulţimea din R, corespunzătoare lui U este

$$I = (-\infty, -u_\alpha) \cup (u_\alpha, +\infty)$$

Testul este

i) dacă $U_0 \in [-u_\alpha, u_\alpha]$, ipoteza H_1 se acceptă

ii) dacă $U_0 \notin [-u_\alpha, u_\alpha]$, ipoteza H_1 se respinge

III. Fie (X,Y) o caracteristică bidimensională urmărită la indivizii unei colectivităţi avînd o repartiţie normală,

cu coeficientul de corelaţie ρ , ce trebuie estimat printr-un interval de încredere.

După cum s-a spus mai înainte, statistica

$$I(X,Y;n,\rho) = \left[Z(X,Y;n) - \frac{1}{2}\ln\frac{1+\rho}{1+\rho}\right]\sqrt{n-3}$$

urmează aproximativ repartiţia $N(0,1)$. Verificăm, pentru statistica $I(X,Y;n,\rho)$, îndeplinirea condiţiilor a), b), c) din §7, cap.IV. Evident, condiţiile a) şi b) sînt îndeplinite. Condiţia c) este îndeplinită întrucît

$$\frac{\partial I(X,Y;n,\rho)}{\partial\rho} = -\frac{1}{1-\rho^2} < 0$$

pentru $\rho\in(-1,1)$.

Pentru nivelul de încredere δ fixat, se poate determina valoarea „a", astfel încît să aibă loc

$$(1)\quad P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ -a \leqslant \left[Z(X,Y;n) - \frac{1}{2}\ln\frac{1+\rho}{1-\rho}\right]\sqrt{n-3} \leqslant a\right) = \delta$$

Relaţia (1) se mai poate scrie sub forma echivalentă

$$(2)\ P^{(n)}\left(\omega^{(n)}\in\Omega^{(n)};\ \frac{e^{2\left[Z(X,Y;n) - \frac{a}{\sqrt{n-3}}\right]} - 1}{e^{2\left[Z(X,Y;n) - \frac{a}{\sqrt{n-3}}\right]} + 1} \leqslant \rho \leqslant \frac{e^{2\left[Z(X,Y,n) + \frac{a}{\sqrt{n-3}}\right]} - 1}{e^{2\left[Z(X,Y;n) + \frac{a}{\sqrt{n-3}}\right]} + 1}\right) = \delta$$

Prin urmare

$$(3)\quad \left[\frac{e^{2\left[Z(X,Y;n) - \frac{a}{\sqrt{n-3}}\right]} - 1}{e^{2\left[Z(X,Y;n) - \frac{a}{\sqrt{n-3}}\right]} + 1},\ \frac{e^{2\left[Z(X,Y;n) + \frac{a}{\sqrt{n-3}}\right]} - 1}{e^{2\left[Z(X,Y;n) + \frac{a}{\sqrt{n-3}}\right]} + 1}\right]$$

este intervalul de încredere de lungime minimă pentru coeficientul de corelaţie a componentelor vectorului aleator bidimensional cu repartiţia normală. Valoarea „a" se determină, cu

ajutorul tabelului IV din anexă, pe baza relaţiei $\Phi(a) = \frac{\delta}{2}$. Ţinînd seama de (3), intervalul determinist corespunzător selecţiei (x_i, y_i), $1 \leq i \leq n$, este

$$\left[\frac{e^{2[z_0 - \frac{a}{\sqrt{n-3}}]} - 1}{e^{2[z_0 - \frac{a}{\sqrt{n-3}}]} + 1}, \frac{e^{2[z_0 + \frac{a}{\sqrt{n-3}}]} - 1}{e^{2[z_0 + \frac{a}{\sqrt{n-3}}]} + 1} \right]$$

unde

$$z_0 = \frac{1}{2} \ln \frac{1 + \bar{r}}{1 - \bar{r}} .$$

§ 8. REGRESIA LINIARĂ DE SELECŢIE

Fie (X,Y) o caracteristică bidimensională urmărită la indivizii unei colectivităţi şi (X_i, Y_i), $1 \leq i \leq n$, vectorii aleatori de selecţie. Considerăm statisticile

$$m(X,n) = \frac{1}{n} \sum_{i=1}^{n} X_i \quad , \quad m(Y,n) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$$d_*^2(X,n) = \frac{1}{n-1} \sum_{i=1}^{n} [X_i - m(X,n)]^2 \quad , \quad d_*^2(Y,n) = \frac{1}{n-1} \sum_{i=1}^{n} [Y_i - m(Y,n)]^2$$

Teorema 1. Minimul expresiei

$$\delta(\alpha,\beta) = \sum_{i=1}^{n} [Y_i - \alpha X_i - \beta]^2$$

se atinge pentru variabilele aleatoare

$$\alpha = r(X,Y;n) \sqrt{\frac{d_*^2(Y,n)}{d_*^2(X,n)}}$$

$$\beta = m(U,n) - r(X,Y;n) \sqrt{\frac{d_*^2(Y,n)}{d_*^2(X,n)}}\, m(X,n)$$

Scanned with OKEN Scanner

Demonstrație. Anulînd derivatele parțiale în raport cu α și β ale lui $S(\alpha,\beta)$, se obține sistemul de ecuații

$$(1)\quad \begin{cases} \dfrac{\partial S(\alpha,\beta)}{\partial \alpha} = -2\sum_{i=1}^{n} X_i(Y_i - \alpha X_i - \beta) = 0 \\ \dfrac{\partial S(\alpha,\beta)}{\partial \beta} = -2\sum_{i=1}^{n} (Y_i - \alpha X_i - \beta) = 0 \end{cases}$$

ceea ce este echivalent cu

$$(2)\quad \begin{cases} (\sum_{i=1}^{n} X_i^2)\alpha + (\sum_{i=1}^{n} X_i)\beta = \sum_{i=1}^{n} X_i Y_i \\ (\sum_{i=1}^{n} X_i)\alpha + n\beta = \sum_{i=1}^{n} Y_i \end{cases}$$

Rezolvînd sistemul de ecuații (2), în raport cu α și β, se obține

$$(3)\quad \alpha = r(X,Y;n)\sqrt{\frac{d_*^2(Y,n)}{d_*^2(X,n)}}$$

$$\beta = m(Y,n) - r(X,Y;n)\sqrt{\frac{d_*^2(Y,n)}{d_*^2(X,n)}}\, m(X,n)$$

Cu aceasta teorema este demonstrată.

Analog se poate demonstra

Teorema 2. Minimul expresiei

$$S'(\alpha',\beta') = \sum_{i=1}^{n} [X_i - \alpha' Y_i - \beta']^2$$

se atinge pentru variabilele aleatoare

$$\alpha' = r(X,Y;n)\sqrt{\frac{d_*^2(X,n)}{d_*^2(Y,n)}}$$

$$\beta' = m(X,n) - r(X,Y;n)\sqrt{\frac{d_*^2(X,n)}{d_*^2(Y,n)}}\, m(Y,n)$$

Analog cu ecuațiile (10) și (13) din §5, cap.VII, ecuația dreptei de regresie de selecție a lui Y în raport cu X este

$$y - m(Y,n) = \alpha\left[x - m(X,n)\right]$$

şi <u>ecuaţia dreptei de regresie de selecţie a lui X în raport cu Y</u> este

$$x - m(X,n) = \alpha'\left[y - m(Y,n)\right]$$

Variabilele aleatoare α şi α' se numesc <u>coeficienţi de regresie liniară de selecţie</u>. Dacă $r(X,Y;n) = \pm 1$, atunci dreptele de regresie de selecţie se confundă şi toate punctele de coordonate (X_i,Y_i), $1 \leq i \leq n$, se află pe dreapta de ecuaţie

$$y - m(Y,n) = \pm\sqrt{\frac{d_*^2(Y,n)}{d_*^2(X,n)}}\left[x - m(X,n)\right]$$

<u>Observaţie.</u> Coeficienţii de regresie liniară de selecţie α şi α' sînt funcţii de estimaţie consistentă pentru coeficienţii de regresie liniară teoretică a şi a'.

P R O B L E M E

1. Repartiţia probabilistică a vectorului aleator (X,Y) este dată de tabelul 16

T a b e l u l 16

X \ Y	-1	1
-1	0,11	0,07
0	0,29	0,19
1	0,20	0,14

Se cere :

a) mediile şi dispersiile marginale

b) probabilităţile condiţionate

c) graficele curbelor de regresie

Scanned with OKEN Scanner

d) coeficientul de contingenţă a variabilelor aleatoare X şi Y.

2. Se dă funcţia

$$f(x,y) = \frac{k}{\sqrt{xy}}, \quad (x,y) \in D$$

unde $k \in R$, iar D este triunghiul determinat de punctele $O(0,0)$, $A(0,1)$, $B(1,1)$.

Se cere :

a) să se determine k încît $f(x,y)$ să fie densitatea de repartiţie a vectorului aleator (X,Y).

b) densităţile de repartiţie marginale şi mediile marginale

c) densităţile de repartiţie condiţionate

d) ecuaţiile curbelor de regresie

e) coeficientul de corelaţie a variabilelor aleatoare X şi Y

3. Se dă funcţia

$$f(x,y) = \begin{cases} k & \frac{x^2}{9} + \frac{y^2}{4} \leq 1 \\ 0 & \frac{x^2}{9} + \frac{y^2}{4} > 1 \end{cases}$$

unde $k \in R$.

Se cere :

a) să se determine k încît $f(x,y)$ să fie densitatea de repartiţie a vectorului aleator (X,Y).

b) probabilitatea ca punctul aleator (X,Y) să aparţină domeniului $D : 0 \leq x \leq \sqrt{3}$, $0 \leq y \leq 1$.

c) mediile şi dispersiile marginale

d) densităţile de repartiţie condiţionate

e) ecuaţiile curbelor de regresie

f) coeficientul de corelaţie a variabilelor aleatoare X şi Y.

4. O repartiţie statistică a unei caracteristici bidimensionale (X,Y) este dată de tabloul de corelaţie (tabelul 17)

Scanned with OKEN Scanner

Tabelul 17

X \ Y	1	2	3	4	5
1	2				3
2	1	1	2		
3		2	2	1	4
4				3	2

Se cere :

a) mediile şi dispersiile marginale statistice

b) mediile condiţionate statistice

c) graficele curbelor de regresie statistică

d) coeficientul de corelaţie statistic

5. Fie X şi Y , abaterile diametrului interior, respectiv exterior, al unor inele, de la dimensiunile standard. Pe baza unor experienţe s-a constatat că, caracteristica (X,Y) urmează o repartiţie normală. In urma cercetării a 20 inele s-au obţinut datele din tabloul de corelaţie (tabelul 18)

Tabelul 18

X \ Y	-2	-1	0
-2	1	3	
-1	2	4	
0		1	5
1		2	2

Scanned with OKEN Scanner

Se cere :

a) coeficientul de corelaţie statistic

b) să se estimeze coeficientul de corelaţie teoretic printr-un interval de încredere, cu nivelul de încredere $\delta = 0,95$

c) să se testeze, la nivelul de semnificaţie $\alpha = 0,05$, ipoteza că, coeficientul de corelaţie teoretic este egal cu 0,8.

Scanned with OKEN Scanner

A N E X Å

T A B E L U L I

Valorile lui $e^{-\lambda}$

λ	$e^{-\lambda}$	λ	$e^{-\lambda}$	λ	$e^{-\lambda}$
0,00	1,000	0,50	0,607	1,0	0,368
01	0,990	51	600	1	333
02	980	52	595	2	301
03	970	53	589	3	273
04	961	54	583	4	247
05	951	55	577	5	223
06	942	56	571	6	202
07	932	57	566	7	183
08	923	58	560	8	165
09	914	59	554	9	150
10	905	60	549	2,0	135
11	896	61	543	1	122
12	887	62	538	2	111
13	878	63	533	3	100
14	869	64	527	4	090 7
15	861	65	522	5	082 1
16	852	66	517	6	074 3
17	844	67	512	7	067 2
18	835	68	507	8	060 8
19	827	69	502	9	055 0
20	819	70	497	3,0	049 8
21	811	71	492	1	045 0
22	803	72	487	2	040 8
23	795	73	482	3	036 9
24	787	74	477	4	033 4
25	779	75	472	5	030 2
26	771	76	468	6	027 3
27	763	77	463	7	024 7
28	756	78	458	8	022 4
29	748	79	454	9	020 2
30	741	80	449	4,0	018 3
31	733	81	445	1	016 6
32	726	82	440	2	015 0
33	719	83	436	3	013 6
34	712	84	432	4	012 3
35	705	85	427	5	011 1
36	698	86	423		
37	691	87	419	5,0	006 74
38	684	88	415	6,0	002 48
39	677	89	411	7,0	000 912
40	670	90	407	8,0	000 335
41	664	91	403	9,0	000 123
42	657	92	399	10,0	000 045
43	651	93	395		
44	644	94	391		
45	638	95	387		
46	631	96	383		
47	625	97	379		
48	619	98	375		
49	613	99	372		
50	607	1,00	368		

Scanned with OKEN Scanner

T A B E L U L II

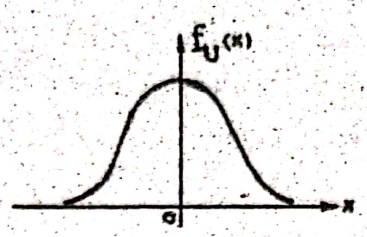
Repartiţia Poisson. Valorile probabilităţii $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

k \ λ	0,2	0,4	0,6	0,8	1	1,5	2	2,5	3	3,5	4	4,5	5	10
0	0,818 7	0,670 3	0,548 8	0,449 3	0,367 9	0,223 1	0,135 3	0,082 1	0,049 8	0,030 2	0,018 3	0,011 1	0,006 7	0,000 0
1	0,163 7	0,268 1	0,329 3	0,359 5	0,367 9	0,334 7	0,270 7	0,205 2	0,149 4	0,105 7	0,370 3	0,050 0	0,033 7	0,000 5
2	0,016 4	0,053 6	0,098 8	0,143 8	0,183 9	0,251 0	0,270 7	0,256 5	0,224 0	0,185 0	0,146 5	0,112 5	0,084 2	0,002 3
3	0,001 1	0,007 1	0,019 8	0,038 3	0,061 3	0,125 5	0,180 4	0,313 8	0,224 0	0,215 8	0,195 4	0,168 7	0,140 4	0,007 6
4		0,000 7	0,003 0	0,007 7	0,015 3	0,047 1	0,090 2	0,133 6	0,168 0	0,188 8	0,195 4	0,189 8	0,175 5	0,018 9
5		0,000 1	0,000 4	0,001 2	0,003 1	0,014 1	0, 1	0,066 8	0,100 8	0,132 2	0,156 3	0,170 8	0,175 5	0,037 8
6				0,000 2	0,000 5	0,003 5	0,012 0	0,02 3	0,050 4	0,077 1	0,104 2	0,128 1	0,146 2	0,063 1
7					0,000 1	0,000 8	0,003 4	0,009 9	0,021 6	0,038 5	0,059 5	0,082 4	0,104 4	0,090 1
8						0,000 1	0,000 9	0,003 1	0,008 1	0,016 9	0,029 8	0,046 3	0,065 3	0,112 6
9							0,000 2	0,000 9	0,002 7	0,006 6	0,013 2	0,023 2	0,036 3	0,125 1
10								0,000 2	0,000 8	0,002 3	0,005 3	0,010 4	0,018 1	0,125 1
11									0,000 2	0,000 7	0,001 9	0,004 3	0,008 2	0,113 7
12									0,000 1	0,000 2	0,000 6	0,001 6	0,003 4	0,094 8
13										0,000 1	0,000 2	0,000 6	0,001 3	0,072 9
14											0,000 1	0,000 2	0,000 5	0,052 1
15												0,000 1	0,000 2	0,034 7
16													0,000 1	0,021 7
17														0,012 8

Scanned with OKEN Scanner

T A B E L U L III

Repartiția N(0,1). Valorile densității de repartiție $f_U(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

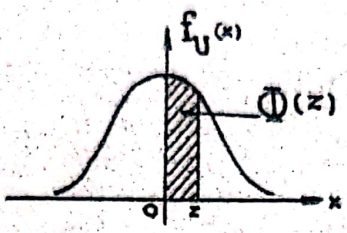


x	$f_U(x)$	x	$f_U(x)$
0,0	0,398 94	2,5	0,017 53
0,1	0,396 95	2,6	0,013 58
0,2	0,391 04	2,7	0,010 42
0,3	0,381 39	2,8	0,007 92
0,4	0,368 27	2,9	0,005 95
0,5	0,352 07	3,0	0,004 43
0,6	0,333 22	3,1	0,003 27
0,7	0,312 25	3,2	0,002 38
0,8	0,289 69	3,3	0,001 72
0,9	0,266 09	3,4	0,001 23
1,0	0,241 97	3,5	0,000 87
1,1	0,217 85	3,6	0,000 61
1,2	0,194 19	3,7	0,000 42
1,3	0,171 37	3,8	0,000 29
1,4	0,149 73	3,9	0,000 20
1,5	0,129 52	4,0	0,000 13
1,6	0,110 92	4,1	0,000 09
1,7	0,094 05	4,2	0,000 06
1,8	0,078 95	4,3	0,000 04
1,9	0,065 62	4,4	0,000 02
2,0	0,053 99	4,5	0,000 02
2,1	0,043 98	4,6	0,000 01
2,2	0,035 47	4,7	0,000 01
2,3	0,028 33	4,8	0,000 00
2,4	0,022 39		

Scanned with OKEN Scanner

T A B E L U L IV

Valorile funcției lui Laplace $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx$



z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
0,00	0,00000	0,40	0,15542	0,80	0,28814	1,20	0,38493
0,01	0,00399	0,41	0,15910	0,81	0,29103	1,21	0,38686
0,02	0,00798	0,42	0,16276	0,82	0,29389	1,22	0,38877
0,03	0,01197	0,43	0,16640	0,83	0,29673	1,23	0,39065
0,04	0,01595	0,44	0,17003	0,84	0,29955	1,24	0,39251
0,05	0,01994	0,45	0,17364	0,85	0,30234	1,25	0,39435
0,06	0,02392	0,46	0,17724	0,86	0,30511	1,26	0,39617
0,07	0,02790	0,47	0,18082	0,87	0,30785	1,27	0,39796
0,08	0,03188	0,48	0,18439	0,88	0,31057	1,28	0,39973
0,09	0,03586	0,49	0,18793	0,89	0,31327	1,29	0,40147
0,10	0,03983	0,50	0,19146	0,90	0,31594	1,30	0,40320
0,11	0,04380	0,51	0,19497	0,91	0,31859	1,31	0,40490
0,12	0,04776	0,52	0,19847	0,92	0,32121	1,32	0,40658
0,13	0,05172	0,53	0,20194	0,93	0,32381	1,33	0,40824
0,14	0,05567	0,54	0,20540	0,94	0,32639	1,34	0,40988
0,15	0,05962	0,55	0,20884	0,95	0,32894	1,35	0,41149
0,16	0,06356	0,55	0,21226	0,96	0,33147	1,36	0,41309
0,17	0,06749	0,57	0,21566	0,97	0,33398	1,37	0,41466
0,18	0,07142	0,58	0,21904	0,98	0,33646	1,38	0,41621
0,19	0,07535	0,59	0,22240	0,99	0,33891	1,39	0,41774
0,20	0,07926	0,60	0,22575	1,00	0,34134	1,40	0,41924
0,21	0,08317	0,61	0,22907	1,01	0,34375	1,41	0,42073
0,22	0,08706	0,62	0,23237	1,02	0,34614	1,42	0,42220
0,23	0,09095	0,63	0,23565	1,03	0,34850	1,43	0,42364
0,24	0,09483	0,64	0,23891	1,04	0,35083	1,44	0,42507
0,25	0,09871	0,65	0,24215	1,05	0,35314	1,45	0,42647
0,26	0,10257	0,66	0,24537	1,06	0,35543	1,46	0,42786
0,27	0,10642	0,67	0,24857	1,07	0,35769	1,47	0,42922
0,28	0,11026	0,68	0,25175	1,08	0,35993	1,48	0,43056
0,29	0,11409	0,69	0,25490	1,09	0,36214	1,49	0,43189
0,30	0,11791	0,70	0,25804	1,10	0,36433	1,50	0,43319
0,31	0,12172	0,71	0,26115	1,11	0,36650	1,51	0,43448
0,32	0,12552	0,72	0,26424	1,12	0,36864	1,52	0,43574
0,33	0,12930	0,73	0,26730	1,13	0,37076	1,53	0,43699
0,34	0,13307	0,74	0,27035	1,14	0,37286	1,54	0,43822
0,35	0,13683	0,75	0,27337	1,15	0,37493	1,55	0,43943
0,36	0,14058	0,76	0,27637	1,16	0,37698	1,56	0,44062
0,37	0,14431	0,77	0,27935	1,17	0,37900	1,57	0,44179
0,38	0,14803	0,78	0,28230	1,18	0,38100	1,58	0,44295
0,39	0,15173	0,79	0,28524	1,19	0,38298	1,59	0,44408

Scanned with OKEN Scanner

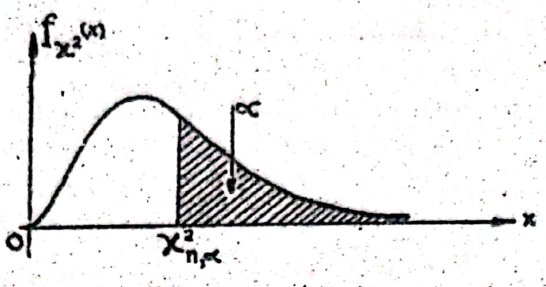
z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
1,60	0,44520	2,03	0,47882	2,46	0,49305	2,89	0,49807
1,61	0,44630	2,04	0,47932	2,47	0,49324	2,90	0,49813
1,62	0,44738	2,05	0,47982	2,48	0,49343	2,91	0,49819
1,63	0,44845	2,06	0,48030	2,49	0,49361	2,92	0,49825
1,64	0,44950	2,07	0,48077	2,50	0,49379	2,93	0,49831
1,65	0,45053	2,08	0,48124	2,51	0,49396	2,94	0,49836
1,66	0,45154	2,09	0,48169	2,52	0,49413	2,95	0,49841
1,67	0,45254	2,10	0,48214	2,53	0,49430	2,96	0,49846
1,68	0,45352	2,11	0,48257	2,54	0,49446	2,97	0,49851
1,69	0,45440	2,12	0,48300	2,55	0,49461	2,98	0,49856
1,70	0,45543	2,13	0,48341	2,56	0,49477	2,99	0,49861
1,71	0,45637	2,14	0,48382	2,57	0,49492	3,00	0,49865
1,72	0,45728	2,15	0,48422	2,58	0,49506	3,01	0,49869
1,73	0,15818	2,16	0,48461	2,59	0,49520	3,02	0,49874
1,74	0,46997	2,17	0,48500	2,60	0,49534	3,03	0,49878
1,75	0,45994	2,18	0,48537	2,61	0,49547	3,04	0,49882
1,76	0,46080	2,19	0,48574	2,62	4,09560	3,05	0,49886
1,77	0,46164	2,20	0,48610	2,63	0,49573	3,06	0,49889
1,78	0,46246	2,21	0,48645	2,64	0,49585	3,07	0,49893
1,79	0,46327	2,22	0,48679	2,65	0,49598	3,08	0,49897
1,80	0,46407	2,23	0,48713	2,66	0,49609	3,09	0,49900
1,81	0,46185	2,24	0,48745	2,67	0,49621	3,10	0,49903
1,82	0,46562	2,25	0,48778	2,68	0,49632	3,11	0,49906
1,83	0,46638	2,26	0,48809	2,69	0,49643	3,12	0,49909
1,84	0,46712	2,27	0,48840	2,70	0,49653	3,13	0,49913
1,85	0,46784	2,28	0,48870	2,71	0,49664	3,14	0,49916
1,86	0,46856	2,29	0,48899	2,72	0,49674	3,15	0,49918
1,87	0,46926	2,30	0,48928	2,73	0,49688	3,16	0,49921
1,88	0,46995	2,31	0,48956	2,74	0,49693	3,17	0,49924
1,89	0,47062	2,32	0,48983	2,75	0,49702	3,18	0,49926
1,90	0,47128	2,33	0,49010	2,76	0,49711	3,19	0,49929
1,91	0,47193	2,34	0,49036	2,77	0,49720	3,20	0,49941
1,92	0,47257	2,35	0,49061	2,78	0,49728	3,21	0,49934
1,93	0,47320	2,36	0,49806	2,79	0,49736	3,22	0,49936
1,94	0,47381	2,37	0,49111	2,80	0,49744	3,23	0,49938
1,95	0,47441	2,38	0,49134	2,81	0,49752	3,24	0,49940
1,96	0,47500	2,39	0,49158	2,82	0,49760	3,25	0,49942
1,97	0,47558	2,40	0,49180	2,83	0,49767	3,26	0,49944
1,98	0,47615	2,41	0,49202	2,84	0,49774	3,27	0,49946
1,99	0,47670	2,42	0,49224	2,85	0,49781	3,28	0,49948
00	0,47725	2,43	0,49245	2,86	0,49788	3,29	0,49950
01	0,47778	2,44	0,49266	2,87	0,49795	3,30	0,49952
02	0,47831	2,45	0,49286	2,88	0,49801	3,31	0,49953

z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
3,32	0,49955	3,60	0,499841	3,88	0,499948	4,16	0,499984
3,33	0,49957	3,61	0,499847	3,89	0,499950	4,17	0,499985
3,34	0,49958	3,62	0,499853	3,90	0,499952	4,18	0,499985
3,35	0,49960	3,63	0,499858	3,91	0,499954	4,19	0,499986
3,36	0,49961	3,64	0,499864	3,92	0,499956	4,20	0,499987
3,37	0,49962	3,65	0,499869	3,93	0,499958	4,21	0,499987
3,38	0,49964	3,66	0,499874	3,94	0,499959	4,22	0,499987
3,39	0,49965	3,67	0,499879	3,95	0,499961	4,23	0,499988
3,40	0,49966	3,68	0,499883	3,96	0,499963	4,24	0,499988
3,41	0,49968	3,69	0,499888	3,97	0,499964	4,25	0,499989
3,42	0,49969	3,70	0,499892	3,98	0,499966	4,26	0,499989
3,43	0,49970	3,71	0,499896	3,99	0,499967	4,27	0,499990
3,44	0,49971	3,72	0,499900	4,00	0,499968	4,28	0,499991
3,45	0,49972	3,73	0,499904	4,01	0,499970	4,29	0,499991
3,46	0,49973	3,74	0,499908	4,02	0,499971	4,30	0,499992
3,47	0,49974	3,75	0,499912	4,03	0,499972	4,31	0,499992
3,48	0,49975	3,76	0,499915	4,04	0,499973	4,32	0,499992
3,49	0,49976	3,77	0,499918	4,05	0,499974	4,33	0,499993
3,50	0,49977	3,78	0,499922	4,06	0,499976	4,34	0,499993
3,51	0,49978	3,79	0,499925	4,07	0,499977	4,35	0,499993
3,52	0,499784	3,80	0,499927	4,08	0,499978	4,36	0,499994
3,53	0,499792	3,81	0,499931	4,09	0,499978	4,37	0,499994
3,54	0,499799	3,82	0,499933	4,10	0,499979	4,38	0,499994
3,55	0,499807	3,83	0,499938	4,11	0,499980	4,39	0,499994
3,56	0,499814	3,84	0,499936	4,12	0,499981	4,40	0,499995
3,57	0,499821	3,85	0,499941	4,13	0,499982	4,45	0,499996
3,58	0,499828	3,86	0,499943	4,14	0,499983	4,50	0,499997
3,59	0,499834	3,87	0,499946	4,15	0,499983	4,60	0,499998
						5,00	0,4999997

Scanned with OKEN Scanner

T A B E L U L V

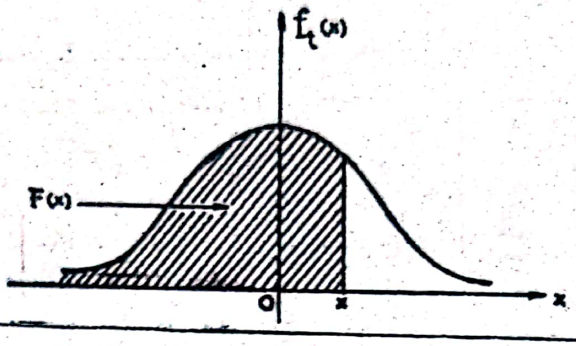
Repartiția $\chi^2(n,1)$. Valorile $\chi^2_{n,\alpha}$ în funcție de probabilitatea $\alpha = P(X > \chi^2_{n,\alpha})$ și numărul gradelor de libertate n .



n \ α	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,662	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,604	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,013	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,243	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

T A B E L U L VI

Repartiţia Student. Valorile funcţiei de repartiţiei F(x) = = P(X < x).

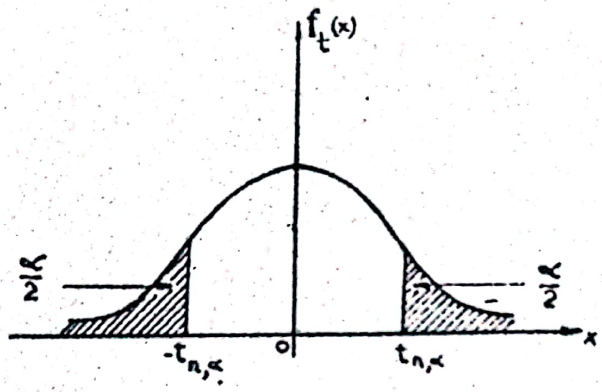


x	n									
	1	2	3	4	5	6	7	8	9	10
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
0,1	532	535	537	537	538	538	538	539	539	539
0,2	563	570	573	574	575	576	576	577	577	577
0,3	593	604	608	610	612	613	614	614	614	615
0,4	621	636	642	645	647	648	650	650	651	651
0,5	648	667	674	678	681	683	684	685	68	686
0,6	672	695	705	710	713	715	716	717	718	719
0,7	694	722	733	739	742	745	747	748	749	750
0,8	715	746	759	766	770	773	775	777	778	779
0,9	733	768	783	790	795	799	801	803	804	805
1,0	0,750	0,789	0,801	0,813	0,818	0,822	0,825	0,827	0,828	0,830
1,1	765	807	824	831	839	843	846	848	850	851
1,2	779	824	842	852	858	862	865	868	870	871
1,3	791	838	858	868	875	879	883	885	887	889
1,4	803	852	872	881	890	894	898	900	902	904
1,5	813	864	885	896	903	908	911	914	916	918
1,6	822	875	896	908	915	920	923	926	928	930
1,7	831	884	906	918	925	930	934	936	938	940
1,8	840	893	915	927	934	939	943	945	947	949
1,9	846	901	923	935	942	947	950	953	955	957
2,0	0,852	0,908	0,930	0,942	0,949	0,954	0,957	0,960	0,962	0,963
2,1	861	921	942	954	960	965	968	970	972	974
2,2	871	931	952	963	969	973	976	978	980	981
2,3	883	938	960	970	976	980	982	984	986	987
2,4	891	946	966	976	981	984	987	988	990	991
3,0	0,898	0,952	0,971	0,980	0,985	0,988	0,990	0,992	0,992	0,993
3,2	908	957	975	984	988	991	992	994	995	995
3,4	909	962	979	986	990	993	994	995	996	997
3,6	914	965	982	989	992	994	996	996	997	998
3,8	918	969	984	990	994	996	997	997	998	998
4,0	0,922	0,971	0,986	0,992	0,995	0,996	0,997	0,998	0,998	0,999
4,2	926	974	988	993	996	997	998	998	999	999
4,4	929	976	989	994	996	998	998	999	999	999
4,6	932	978	990	995	997	998	999	999	999	1,000
4,8	935	980	991	996	998	998	999	999	1,000	
5,0	0,937	0,981	0,992	0,996	0,998	0,999	0,999	1,000		
5,2	940	982	993	997	998	999	999			
5,4	942	984	994	997	998	999	1,000			
5,6	944	985	994	998	999	999				
5,8	946	986	995	998	999	999				
6,0	0,947	0,987	0,995	0,998	0,999	1,000				

n										
11	12	13	14	15	16	17	18	19	∞	x
0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,50000	0,0
539	539	539	539	539	539	539	539	539	53983	0,1
577	578	578	578	578	578	578	578	578	57926	0,2
615	615	616	616	616	616	616	616	616	61791	0,3
652	652	652	652	653	653	653	653	653	65542	0,4
686	687	687	688	688	688	688	688	689	69146	0,5
720	720	721	721	721	722	722	722	722	72575	0,6
751	751	752	752	753	753	753	754	754	75804	0,7
780	780	781	782	782	782	783	783	783	78814	0,8
806	807	808	808	809	809	810	810	810	81594	0,9
0,831	0,832	0,832	0,833	0,833	0,834	0,834	0,835	0,835	0,84134	1,0
853	854	854	855	856	856	857	857	858	86433	1,1
872	873	874	875	876	876	877	877	878	88493	1,2
890	891	892	893	893	894	894	895	895	90320	1,3
906	907	908	908	909	910	910	911	911	91924	1,4
919	920	921	922	923	924	924	924	925	93319	1,5
931	932	933	934	935	935	936	936	937	94520	1,6
941	943	944	945	945	946	946	947	947	95543	1,7
950	952	952	953	954	955	955	956	956	96407	1,8
958	959	960	961	962	962	963	963	964	97128	1,9
0,965	0,967	0,967	0,967	0,968	0,969	0,969	0,970	0,970	0,97725	2,0
975	976	977	977	978	979	979	979	980	98610	2,2
982	983	984	985	985	986	986	986	987	99180	2,4
988	988	989	990	990	990	991	991	991	99534	2,6
991	992	992	993	993	994	994	994	994	99744	2,8
0,994	0,994	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,99865	3,0
996	996	996	997	997	997	997	998	998	99931	3,2
997	997	998	998	998	998	998	998	998	99966	3,4
998	998	998	999	999	999	999	999	999	99984	3,6
998	999	999	999	999	999	999	999	999	99993	3,8
0,999	0,999	0,999	0,999	0,999	1,000	1,000	1,000	1,000	0,99997	4,0
999	999	1,000	1,000	1,000					99999	4,2
1,000	1,000								99999	4,4

T A B E L U L VII

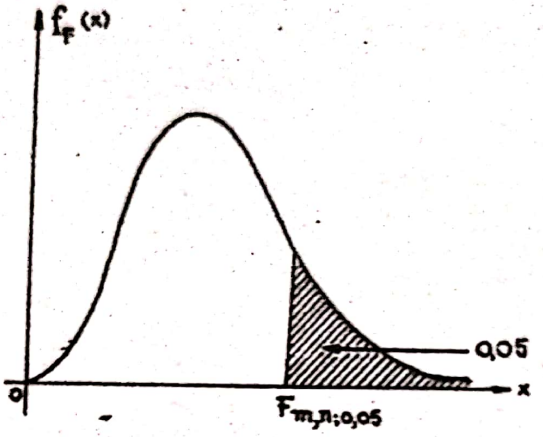
Repartiţia Studont. Valorile $t_{n,\alpha}$ în funcţie de probabilitatea $\alpha = P(|X| > t_{n,\alpha})$ şi numărul gradelor de libertate n.



n \ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
	0,125 66	0,253 35	0,385 32	0,524 40	0,674 49	0,841 62	1,036 43	1,281 55	1,644 85	1,959 96	2,326 34	2,575 82

T A B E L U L VIII

Repartiţia Snedecor. Valorile $F_{m,n;0,05}$ în funcţie de probabilitatea $\alpha = P(X > F_{m,n;0,05}) = 0,05$ şi numerele gradelor de libertate m şi n

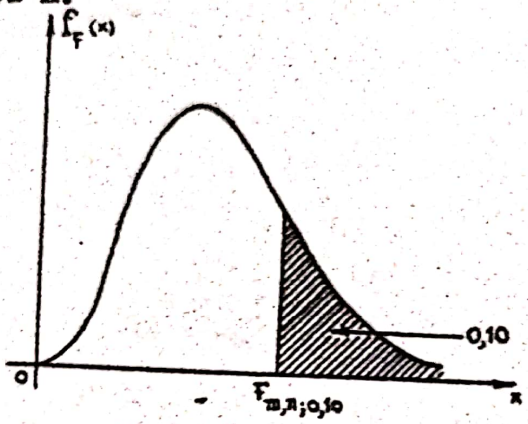


n \ m	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,552	9,277	9,117	9,014	8,941	8,887	8,845	8,812
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999
5	6,608	5,786	5,410	5,192	5,050	4,950	4,876	4,818	4,773
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677
8	5,318	4,459	4,066	3,838	3,688	3,581	3,501	3,438	3,388
9	5,117	4,257	3,863	3,633	3,482	3,374	3,293	3,230	3,179
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646
15	4,543	3,682	3,287	3,056	2,901	2,791	2,707	2,641	2,588
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393
21	4,325	3,467	3,073	2,840	2,685	2,573	2,488	2,421	2,366
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265
27	4,210	3,354	2,960	2,728	2,572	2,459	2,374	2,305	2,250
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211
40	4,085	3,232	2,839	2,606	2,450	2,336	2,249	2,180	2,124
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880

n	m									
	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.09	251.14	252.20	253.25	254.32
2	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	8.786	8.745	8.703	8.660	8.639	8.617	8.594	8.572	8.549	8.527
4	5.964	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.628
5	4.735	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.398	4.365
6	4.060	4.000	3.938	3.874	3.842	3.808	3.774	3.740	3.705	3.669
7	3.637	3.575	3.511	3.445	3.411	3.376	3.340	3.304	3.267	3.230
8	3.347	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.928
9	3.137	3.073	3.006	2.937	2.901	2.864	2.826	2.787	2.748	2.707
10	2.978	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	2.854	2.788	2.719	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	2.753	2.687	2.617	2.544	2.506	2.466	2.426	2.384	2.341	2.296
13	2.671	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	2.602	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	2.544	2.475	2.404	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	2.494	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	2.450	2.381	2.308	2.230	2.190	2.148	2.104	2.058	2.011	1.960
18	2.412	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	2.378	2.308	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	2.348	2.278	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	2.321	2.250	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	2.297	2.226	2.151	2.071	2.028	1.984	1.938	1.890	1.838	1.783
23	2.275	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	2.255	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	2.237	2.165	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	2.220	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	2.204	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	2.190	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	2.177	2.105	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	2.165	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.684	1.622
40	2.077	2.001	1.925	1.839	1.793	1.744	1.693	1.637	1.577	1.509
60	1.993	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	1.911	1.834	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
∞	1.831	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

T A B E L U L IX

Repartiţia Snedecor. Valorile $F_{m,n;o,1o}$ în funcţie de probabilitatea $\alpha = P(X > F_{m,n;o,1o}) = 0,10$ şi numerele gradelor de libertate m şi n.



n \ m	1	2	3	4	5	6	7	8	9
1	39.864	49,500	53,593	55.833	57.241	58.204	58.906	59,439	59.858
2	8.526	9,000	9,162	9,243	9.293	9.326	9,349	9.367	9.381
3	5.538	5,462	5,391	5,343	5.309	5,285	5.266	5.252	5.240
4	4,545	4,325	4,191	4.107	4,051	4,010	3,979	3,955	3,936
5	4,060	3,781	3,620	3,520	3,453	3,405	3,368	3,339	3,316
6	3,776	3,463	3,289	3,181	3,108	3,055	3,015	2,983	2,958
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725
8	3,458	3,113	2,924	2,806	2,727	2,668	2,624	2,589	2,561
9	3,360	3,007	2,813	2,693	2,611	2,551	2,505	2,469	2,440
10	3,285	2,925	2,728	2,605	2,522	2,461	2,414	2,377	2,347
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164
14	3,102	2,727	2,522	2,395	2,307	2,243	2,193	2,154	2,122
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086
16	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055
17	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028
18	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005
19	2,990	2,606	2,397	2,266	2,176	2,109	2,058	2,017	1,984
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965
21	2,961	2,575	2,365	2,233	2,142	2,075	2,023	1,982	1,948
22	2,949	2,561	2,351	2,219	2,128	2,061	2,008	1,967	1,933
23	2,937	2,549	2,339	2,207	2,115	2,047	1,995	1,953	1,919
24	2,927	2,538	2,327	2,195	2,103	2,035	1,983	1,941	1,906
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895
26	2,909	2,519	2,308	2,175	2,082	2,014	1,961	1,919	1,884
27	2,901	2,511	2,299	2,166	2,073	2,005	1,952	1,909	1,874
28	2,894	2,503	2,291	2,157	2,065	1,996	1,943	1,900	1,865
29	2,887	2,496	2,283	2,149	2,057	1,988	1,935	1,892	1,857
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849
40	2,835	2,440	2,226	2,091	1,997	1,927	1,873	1,829	1,793
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738
120	2,748	2,347	2,130	1,992	1,896	1,824	1,768	1,722	1,684
∞	2,706	2,303	2,084	1,945	1,847	1,774	1,717	1,670	1,632

n \ m	10	12	15	20	24	30	40	60	120	∞
1	60,195	60,705	61,220	61,740	62,002	62,265	62,529	62,794	63,061	63,328
2	9,392	9,408	9,425	9,441	9,450	9,458	9,466	9,475	9,483	9,491
3	5,230	5,216	5,200	5,185	5,176	5,168	5,160	5,151	5,143	5,134
4	3,920	3,896	3,869	3,844	3,831	3,817	3,804	3,790	3,775	3,761
5	3,297	3,268	3,238	3,207	3,191	3,174	3,157	3,140	3,123	3,105
6	2,937	2,905	2,871	2,836	2,818	2,800	2,781	2,762	2,742	2,722
7	2,703	2,668	2,632	2,595	2,575	2,556	2,535	2,514	2,493	2,471
8	2,538	2,502	2,464	2,425	2,404	2,383	2,361	2,339	2,316	2,293
9	2,416	2,379	2,340	2,298	2,277	2,255	2,232	2,209	2,184	2,159
10	2,322	2,284	2,244	2,201	2,178	2,155	2,132	2,107	2,082	2,055
11	2,248	2,209	2,167	2,123	2,100	2,076	2,052	2,026	2,000	1,972
12	2,188	2,147	2,105	2,060	2,036	2,012	1,986	1,960	1,932	1,904
13	2,138	2,097	2,053	2,007	1,983	1,958	1,932	1,904	1,876	1,846
14	2,095	2,054	2,010	1,963	1,938	1,912	1,885	1,857	1,828	1,797
15	2,059	2,017	1,972	1,924	1,899	1,873	1,845	1,817	1,787	1,755
16	2,028	1,985	1,940	1,891	1,866	1,839	1,811	1,782	1,751	1,718
17	2,001	1,958	1,912	1,862	1,836	1,809	1,781	1,751	1,719	1,686
18	1,977	1,933	1,887	1,837	1,810	1,783	1,754	1,723	1,691	1,657
19	1,956	1,912	1,865	1,814	1,787	1,759	1,730	1,699	1,666	1,631
20	1,937	1,892	1,845	1,794	1,767	1,738	1,708	1,677	1,643	1,607
21	1,920	1,875	1,827	1,776	1,748	1,719	1,689	1,657	1,623	1,586
22	1,904	1,859	1,811	1,759	1,731	1,702	1,671	1,639	1,604	1,567
23	1,890	1,845	1,796	1,744	1,716	1,686	1,655	1,622	1,587	1,549
24	1,878	1,832	1,783	1,730	1,702	1,672	1,641	1,607	1,572	1,533
25	1,866	1,820	1,771	1,718	1,689	1,659	1,627	1,593	1,557	1,518
26	1,855	1,809	1,760	1,706	1,677	1,647	1,615	1,581	1,544	1,504
27	1,845	1,799	1,749	1,695	1,666	1,636	1,603	1,569	1,531	1,491
28	1,836	1,790	1,740	1,685	1,655	1,625	1,593	1,558	1,520	1,478
29	1,827	1,781	1,731	1,676	1,647	1,616	1,583	1,547	1,509	1,467
30	1,820	1,773	1,722	1,667	1,638	1,607	1,573	1,538	1,199	1,456
40	1,763	1,715	1,662	1,605	1,574	1,511	1,506	1,467	1,425	1,377
60	1,707	1,657	1,603	1,544	1,511	1,476	1,437	1,395	1,348	1,292
120	1,652	1,601	1,545	1,482	1,447	1,409	1,368	1,320	1,265	1,193
∞	1,599	1,546	1,487	1,421	1,383	1,342	1,295	1,240	1,169	1,000

Scanned with OKEN Scanner

T A B E L U L X

Valorile funcţiei lui Kolmogorov $K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$

λ	K(λ)	λ	K(λ)	λ	K(λ)
0,28	0,000001	0,71	0,305471	1,14	0,851394
0,29	0,000004	0,72	0,322265	1,15	0,858038
0,30	0,000009	0,73	0,339113	1,16	0,864442
0,31	0,000021	0,74	0,355961	1,17	0,870612
0,32	0,000046	0,75	0,372833	1,18	0,876548
0,33	0,000091	0,76	0,389640	1,19	0,882258
0,34	0,000171	0,77	0,406372	1,20	0,887750
0,35	0,000303	8,78	0,423002	1,21	0,893030
0,36	0,000511	0,79	0,439505	1,22	0,898104
0,37	0,000826	0,80	0,455857	1,23	0,902972
0,38	0,001285	0,81	0,472041	1,24	0,907648
0,39	0,001929	0,82	0,488030	1,25	0,912132
0,40	0,002808	0,83	0,503808	1,26	0,916432
0,41	0,003972	0,84	0,519366	1,27	0,920556
0,42	0,005476	0,85	0,534682	1,28	0,924505
0,43	0,007377	0,86	0,549744	1,29	0,928288
0,44	0,009430	0,87	0,564548	1,30	0,931908
0,45	0,012590	0,88	0,579070	1,31	0,935370
0,46	0,016005	0,89	0,593316	1,32	0,938682
0,47	0,020022	0,90	0,607270	1,33	0,941848
0,48	0,024682	0,91	0,620928	1,34	0,944872
0,49	0,030017	0,92	0,634286	1,35	0,947756
0,50	0,036055	0,93	0,647338	1,36	0,950512
0,51	0,042814	0,94	0,660082	1,37	0,953142
0,52	0,050306	0,95	0,672516	1,38	0,955650
0,53	0,058534	0,96	0,684636	1,39	0,958040
0,54	0,067497	0,97	0,696444	1,40	0,960318
0,55	0,077183	0,98	0,707940	1,41	0,962486
0,56	0,087577	0,99	0,719126	1,42	0,964552
0,57	0,098656	1,00	0,730000	1,43	0,966516
0,58	0,110395	1,01	0,740566	1,44	0,968382
0,59	0,122760	1,02	0,750826	1,45	0,970158
0,60	0,135718	1,03	0,760780	1,46	0,971846
0,61	0,149229	1,04	0,770434	1,47	0,973448
0,62	0,163225	1,05	0,779794	1,48	0,974970
0,63	0,177753	1,06	0,788860	1,49	0,976412
0,64	0,192077	1,07	0,797636	1,50	0,977782
0,65	0,207987	1,08	0,806128	1,51	0,979080
0,66	0,223637	1,09	0,814342	1,52	0,980310
0,67	0,239582	1,10	0,822282	1,53	0,981476
0,68	0,255780	1,11	0,829950	1,54	0,982578
0,69	0,272189	1,12	0,837356	1,55	0,983622
0,70	0,288765	1,13	0,844502	1,56	0,984610

Scanned with OKEN Scanner

λ	K(λ)	λ	K(λ)	λ	K(λ)
1,57	0,985544	1,92	0,998745	2,27	0,999934
1,58	0,986426	1,93	0,998837	2,28	0,999940
1,59	0,987260	1,94	0,998924	2,29	0,999944
1,60	0,988048	1,95	0,999001	2,30	0,999949
1,61	0,988791	1,96	0,999079	2,31	0,999954
1,62	0,989492	1,97	0,999149	2,32	0,999958
1,63	0,990154	1,98	0,999213	2,33	0,999962
1,64	0,990777	1,99	0,999273	2,34	0,999965
1,65	0,991364	2,00	0,999329	2,35	0,999968
1,66	0,991917	2,01	0,999380	2,36	0,999970
1,67	0,992438	2,02	0,999428	2,37	0,999973
1,68	0,992928	2,03	0,999474	2,38	0,999976
1,69	0,993389	2,04	0,999516	2,39	0,999978
1,70	0,993823	2,05	0,999552	2,40	0,999980
1,71	0,994230	2,06	0,999588	2,41	0,999982
1,72	0,994612	2,07	0,999620	2,42	0,999984
1,73	0,994971	2,08	0,999650	2,43	0,999986
1,74	0,995309	2,09	0,9	2,44	0,999987
1,75	0,995626	2,10	0,9	2,45	0,999988
1,76	0,995922	2,11	0,999728	2,46	0,999988
1,77	0,996200	2,12	0,999750	2,47	0,999990
1,78	0,996460	2,13	0,999770	2,48	0,999991
1,79	0,996704	2,14	0,999790	2,49	0,999992
1,80	0,996932	2,15	0,999806	2,50	0,9999925
1,81	0,997146	2,16	0,999822	2,55	0,9999956
1,82	0,997346	2,17	0,999838	2,60	0,9999974
1,83	0,997533	2,18	0,999852	2,65	0,9999984
1,84	0,997707	2,19	0,999864	2,70	0,9999990
1,85	0,997870	2,20	0,999874	2,75	0,9999994
1,86	0,998023	2,21	0,999886	2,80	0,9999997
1,87	0,998145	2,22	0,999896	2,85	0,99999982
1,88	0,998297	2,23	0,999904	2,90	0,99999990
1,89	0,998421	2,24	0,999912	2,95	0,99999994
1,90	0,998536	2,25	0,999920	3,00	0,99999997
1,91	0,998644	2,26	0,999926		

B I B L I O G R A F I E

1. Ciucu G. : Elemente de teoria probabilităţilor şi statistică matematică. Bucureşti, Ed. did.ped., 1963.

2. Ciucu G.
Craiu V. : Teoria estimaţiei şi verificarea ipotezelor statistice. Bucureşti, Ed. did. ped., 1968.

3. Ciucu G.
Craiu V. : Introducere în teoria probabilităţilor şi statistică matematică. Bucureşti. Ed.did. ped., 1971.

4. Craiu I.
Mihoc Gh.
Craiu V. : Matematici pentru economişti, vol.2. Bucureşti, Ed. tehnică, 1970.

5. Cramer H. : Matematiceskie metodî statistiki. Moskva, 1948.

6. Ionescu H.M. : Statistică matematică. Bucureşti, Ed.did. ped., 1962

7. Iosifescu M.
Mihoc Gh.
Theodorescu R. : Teoria probabilităţilor şi statistica matematică. Bucureşti, Ed. tehnică,1966.

8. Mihoc Gh.
Urseanu V. : Matematici aplicate în statistică. Bucureşti, Ed. Acad. R.P.R., 1962.

9. Mihoc G.
Firescu D. : Statistică matematică. Bucureşti, Ed. did. ped., 1966.

10. Mihoc Gh.
Ciucu G.
Craiu V. : Teoria probabilităţilor şi statistică matematică. Bucureşti, Ed.did.ped.,1970.

11. Mood. M. : The mathematical statistics, London,1951.

12. Onicescu O. : Lecţii de statistică matematică.Bucureşti,
Mihoc Gh. : Ed. tehnică, 1958.

13. Onicescu O. : Calculul probabilităţilor şi aplicaţii. Bucureşti, Ed. Acad. R.P.R., 1956.
Mihoc G.
Ionescu C.T.-
Tulcea

Scanned with OKEN Scanner

14. Rancu N. : Statistica matematică cu aplicaţii în
Tövissi L. producţie. Bucureşti, Ed. Acad. R.P.R.,
1963.

15. Reischer Corina : Teoria probabilităţilor. Bucureşti, Ed. did.
Sâmboan G. ped., 1967.
Theodorescu R.

16. Wilks S.S. : Matematiceskaia statistika. Moskova, 1967.

- oooOooo -